General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

CR 160100

NAS 9-14629

EARLY DETECTION OF DISEASE: THE CORRELATION OF THE VOLATILE ORGANIC PROFILES FROM PATIENTS WITH UPPER RESPIRATORY INFECTIONS WITH SUBJECTS OF NORMAL PROFILE

(NASA-CR-160100) EARLY DETECTION OF DISEASE: THE CORRELATION OF THE VOLATILE ORGANIC PROFILES FROM PATIENTS WITH UPPER RESPIRATORY INFECTIONS WITH SUBJECTS OF NORMAL PROFILES Final Report (Houston

N79-17532

G3/52

Unclas
13954

Final Report

January, 1979

Principal Investigator: Dr. Albert Zlatkis

Chemistry Department

University of Houston

Houston, Texas 77004

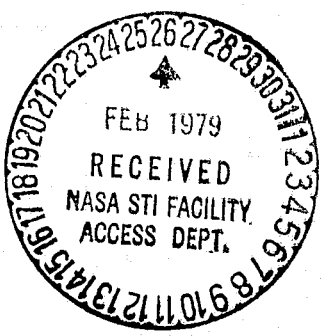


TABLE OF CONTENTS

	PAGE
Abstract	i
I. INTRODUCTION	1
II. EXPERIMENTAL	7
A. Adsorbents and Reagents	7
B. Serum Samples	8
C. Transevaporator Sampling Method	9
1. Sampling Apparatus	9
2. Sampling Procedure	9
3. Transference of Sample from Trapping Tube to Analytical Column	13
D. Capillary Column Gas Chromatography	13
1. Analytical Capillary Column	13
2. Gas Chromatographic Conditions	15
3. Gas Chromatography-Mass Spectrometry	15
E. Data Handling and Computer Interpretation	16
1. Autoscaling	16
2. K-Nearest-Neighbor Classification	16
3. Computer Programs for Profile Analysis	18

TABLE OF CONTENTS

		PAGE
III.	RESULTS AND DISCUSSION	39
	A. Transevaporator Sampling System	39
	B. Volatile Compounds in Virus Infected Serum	50
	C. Profile Analysis of Virus Infected Serum	55
	1. Differentiation between Normal Serum and Virus Infected Serum	61
	2. Differentiation between Normal Serum with Clinical Symptoms after Exposure and Normal Serum with no Clinical Symptoms after Exposure	69
IV.	CONCLUSION	71
V.	BIBLIOGRAPHY	72

LIST OF FIGURES

FIGURE		PAGE
1.	Modified "Transevaporator" Sampling Apparatus	10
2.	Schematic Diagram of a Combined Headspace and Extraction Sampling Procedure	11
3.	Schematic Diagram of Data Processing of Chromatographic Information	17
4.	Chromatogram of a Ten Standard Component Mixture on 5% Witconol Stainless Steel Capillary Column	41
5.	Three Replicate Chromatograms of Volatile Compounds in 60 µl of Normal Serum. Transevaporator used in the Glass Bead Mode	44
6.	Plot of Peak Area vs Concentration of Various Compounds Dissolved in 100 µl of Aqueous Sample	45
7.	Plot Showing Percentage Recovery vs Concentration for Five Compounds Dissolved in 100 µl of Aqueous Sample ...	46
8.	Chromatogram of Volatile Compounds in 60 µl Serum Collected from a Diabetic Patient. Transevaporator Used in the Glass Bead Mode.....	48
9.	Chromatogram of Volatile Compounds in 60 µl Serum Collected from a Diabetic Patient. Transevaporator Used in the Tenax Mode	49
10.	Chromatogram of Volatile Compounds in 200 µl of Saliva. Transevaporator Used in the Glass Bead Mode	51
11.	Chromatogram of Volatile Compounds in 50 µl of Breastmilk. Transevaporator Used in the Glass Bead Mode	52
12.	Chromatogram of Volatile Compounds in 50 µl of Cognac ..... A: Transevaporator Used in the Tenax Mode B: Transevaporator Used in the Glass Bead Mode	53

LIST OF FIGURES (cont.)

FIGURE		PAGE
13.	Chromatogram of Volatile Compounds in 100 µl of Hot Coffee. Transevaporator Used in the Glass Bead Mode	54
14.	Profile of Organic Volatiles from a Pooled Virus Infected Serum by Capillary Column Gas Chromatography	56
15.	Mass Spectra of Peak Number 17 in Virus Infected Serum Chromatogram (Fig. 14), Standard Cyclohexanone and Standard 5-methyl-3-heptanone	59
16.	Autoscaled Two Peak Ratio Diagram ..... I. Normal Serum before Infection; II. Infected sera (24 hr.); o, Recovered sera (14-21 days) After Infection	66
17.	Computer Print Out for Euclidean Distance of 26 Samples Using 10 Known Samples as a Training Set in Two Peak Ratio Factor Space	67
18.	Final Computer Print Out of K-NN Method to Differentiate Before and After Infection Groups	68
19.	Two Peak Ratio Diagram. Sample I1 through I6 are Normal Samples from Volunteers who Contracted Virus Infection After Exposure. I7 through I12, As Above Except no Clinical Symtoms Developed After Exposure	70

LIST OF TABLES

TABLE		PAGE
I.	Standard Deviations of Retention Time and Peak Area	42
II.	Substances Identified by Gas Chromatography-Mass Spectrometry in Virus Infected Serum	57
III.	Reproducibility of Normalized Peak Area and Retention Time Within a Pooled Serum Sample	62
IV.	Variation of Normalized Peak Areas for the Selected Four Peaks in All Serum Samples	63
V.	Selected Two Peak Ratio Data at Different Virus Infected Conditions	64

ABSTRACT

A method is described whereby a transevaporator is used for sampling 60-100 µl of aqueous sample. Volatiles are stripped from the sample either by a stream of helium and collection on a porous polymer, Tenax, or by 0.8 ml of 2-chloropropane and collected on glass beads. The volatiles are thermally desorbed into a precolumn which is connected to a capillary gas chromatographic column for analysis. The technique is shown to be reproducible and suitable for determining chromatographic profiles for a wide variety of sample types.

Using a transevaporator sampling technique, the volatile profiles from 70 µl of serum were obtained by capillary column gas chromatography. The complex chromatograms were interpreted by a combination of manual and computer techniques and a two peak ratio method devised for the classification of normal and virus infected sera. Using the K-Nearest Neighbor approach, 85.7% of the unknown samples were classified correctly. Some preliminary results indicate the possible use of the method for the assessment of virus susceptibility.

I. INTRODUCTION

Profile analysis has been the subject of considerable investigation, with various degrees of success, in food quality evaluation,[1-12] in air pollution studies,[13,78] in the characterization of automobile exhaust gases,[11] marine sediments,[14] engine oils,[15] and marine oil spills.[16-18] The object of gas chromatographic profile analysis is to correlate a characteristic "finger print" pattern in a chromatogram with the qualitative and quantitative properties of a sample.

In recent years, this approach has become increasingly important in the biomedical field for the diagnosis of disease and the study of metabolic disorders.[19-52] A complete profile of all the constituents of a biological fluid is at present impossible with available analytical techniques. The goal at which most workers aim is the development of a complete profile of a selected group of substances (eg. organic acids,[30, 31,36,44,45] amino acids,[32] steroids,[22,39,46] carbohydrates,[40,41] etc.) or of compounds with similar properties (eg. volatiles).[19-21,23,24,28,37,38]

A comparison is then made between the profile obtained with "normal samples" and "pathological samples" to establish any quantitative differences that might be of value for diagnostic purposes. A knowledge of the chemical constitution of abnormal peaks or peaks of abnormal concentration may suggest a possible biochemical reason for the disease and lead to new methods of treatment.

Except in a few cases in which a strong clue exists to implicate the role of causative agents to a disease disorder most work still relies on

a serendipitious finding in a screening program involving the study of many diseases. To maximize the chance of discovery as many substances as possible are included in the profile. Capillary columns are generally used for the analyses due to their superior resolving power.

Volatile metabolites excreted in urine have been extensively studied by Zlatkis et al,[19,47,50] Liebich et al,[21,37,51,52] Matsumoto[48] and Novotny.[49] Abnormal excretion of C_4-C_5 alcohols, octanols, cyclohexanone, 2-heptanone and pyrazines occur in cases of metabolic disorders of diabetes mellitus.

Clinical laboratories depend to a great extent on blood as opposed to urine for diagnostic purposes. Little data is available concerning the occurrence and concentration of volatile substances in blood. Technical problems such as the availability of sufficiently large samples and the protein content of blood serum have rendered their qualitative and quantitative determination more difficult. Dowty et al[53] used Tenax GC to analyze volatiles in plasma of patients with chronic renal failure. Stoner et al[25] extracted volatile substances from 15 ml of plasma with ether. Liebich and Wöll[38] analyzed 5 ml of serum by GC-MS and mass fragmentography using a headspace sampling technique. In all these studies the sample size used for analysis ranged between 5 ml and 55 ml and the chromatograms produced contained relatively few volatile components compared to urine.

The components of interest are generally present in low concentration with the result that special techniques are necessary to overcome the problems associated with the small sample size, low sample capacity of

the capillary column and the need to introduce the sample as a concentrated plug. Two techniques that have been widely used for the concentration of volatile substances from dilute aqueous solution are solvent extraction[8,9,11,14-18,25,32,34,35,41] and headspace analysis using a suitable adsorbent.[9,10,12,13,21,23,37,38,54,55] The techniques are complementary for samples that contain organics covering a wide volatility range. A shortcoming of the extraction technique is that some early peaks are masked by solvent whereas poor representation of late peaks occurs with the headspace technique. The extraction may be performed using a micro-technique but the headspace method generally yields unsatisfactory results where only small volumes of sample are available.

Zlatkis and coworkers introduced a direct micro-extraction method for profile analysis using only 25-100 µl of serum.[56-57] Zlatkis and Kim have recently described an improved method of micro-extraction using a transevaporator.[58] An aqueous sample is placed on a suitable adsorbent packed in a small tube, the organics of interest are extracted with a suitable solvent and transferred to a collection tube packed with glass wool and excess solvent is evaporated. The volatiles are then transferred to a capillary column via thermal desorption into a precolumn. The extraction procedure takes less than 30 minutes and may be used with as little as 25 µl of sample. The transevaporator extraction is about three times more efficient than a direct extraction. Purified 2-chloropropane was found to be a satisfactory extracting solvent and Porasil E was found to be the best adsorbent out of 6 possible candidates.

The volatiles are present at trace levels in biological fluids which consist principally of water so that the sampling technique has to serve as a concentration device capable of reproducibly stripping a diverse range of compounds from a water matrix. The original "transevaporator" method has several shortcomings. A small excess of moisture sometimes causes plugging of the analytical capillary column, and the chromatogram was lacking in headspace components. The recovery and reproducibility of standard compounds were also lower than expected. It was necessary to modify and improve this "transevaporator" sampling system for its general application to biological fluids. Good precision was required and a more general profile containing "headspace" and "solvent extraction" information from the same sample.

When one is dealing with the metabolic disorders that lead to gross metabolic changes, there is no need for computer aided evaluation. However some clinical disorders may result in more subtle, yet perhaps important, changes in the metabolic pattern. The very nature of this approach leads to problems of an analytical nature. For the case of the analysis of volatiles, the current sampling techniques generate very complex chromatograms which only reveal a full picture of their complexity when high resolving capillary columns are used for the gas chromatographic separation. The amount of information subsequently generated can no longer be handled by empirical means and the use of pattern recognition techniques and computer sorting are obligatory.[71] Although the selected sample can be characterized as "volatile" under that heading is masked the chemical complexity of the mixture which covers the

complete spectrum of polarity. The physical property shared by all components is that they fit into a distinct boiling point range. The successful use of such techniques as pattern recognition makes high demands on the reproducibility of the profile; a feature which it is all too easy to demonstrate is more often impaired by poor sampling techniques than chromatographic error.[72]

Various pattern recognition techniques have been described to correlate gas chromatographic profiles with the properties of a sample. The purpose of pattern recognition is generally to categorize observed experimental data as being characteristic of a particular sample type. Moskowitz[3] used multi linear regression analysis to yield an overall dissimilarity of odor quality difference in flavor assessment. Jurs et al[59-62] used a linear learning machine pattern classifier to analyze low resolution mass spectra. The linear learning machine performs as a binary pattern classifier - yes or no, plus or minus, or class I or class II. A classification vector is computed by an interactive training data set using known spectra that may overlap in any or all components. This technique has a disadvantage for the non-linear multi-pattern data set compared to the K-Nearest Neighbor technique (KNN). The K-NN classification method[63-67] is a multi-dimensional classifier with no increase in complexity and is computationally and conceptionally quite simple. An unknown pattern is classified according to the majority vote of its K-Nearest Neighbor (K = 1,3,5...etc.) in n-space. Computationally, all that is necessary is to calculate and scan a distance matrix between an unknown point and points of a known group established from the analysis of a training set.

Profile analysis of volatile compounds in body fluids could be a significant part of a large-scale preventive medicine program provided that an accurate and long-term reproducible analytical system could be developed. Specimen collection, storage and transportation should be controlled carefully in order to eliminate any artifacts. The combination of high resolution capillary column gas chromatography and computer aided pattern classification systems, opens up new possibilities not only for the study of gross alterations but also for the detection of more subtle changes in profiles.

In an effort to establish the methodology for the early diagnosis of viral diseases, a series of capillary column chromatogrpahic profiles were obtained using a modified "transevaporator" sampling technique. The complex chromatographic profiles were analyzed by computer techniques to develop a means of differentiation between normal and virus infected sera using a two peak ratio method. A training set was developed and the K-Nearest Neighbor technique used to establish the predictability of virus infected serum identification. Gas chromatography-mass spectrometry was used to identify the most prominent serum components important to this study.

II. EXPERIMENTAL

A. Adsorbents and Reagents

For trapping and concentrating the volatiles from biological fluids, the adsorbent should provide adequate selectivity, thermal stability and surface activity.

Porasil E (80-100 mesh) and glass beads (80-100 mesh) were obtained from Analabs, North Haven, Connecticut. Porasil E is in the form of perfectly spherical beads of porous silica with a specific surface area of 25 m^2/g. According to Cadogan and Sawyer,[73] upon heating from 200°C to 500°C silica loses all hydrogen-bonded water which provides the active sites for adsorption. Since Porasil E was conditioned at 300°C prior to use, water in serum was retained by Porasil E very effectively after sample elution through this small column.

Tenax GC (60-80 mesh), a porous polymer of 2,6-diphenyl-p-phenylene oxide, was obtained from Applied Science, State College, Pennsylvania. Tenax GC has been characterized as having unusual and desirable chromatographic properties.[74] It has been shown to fulfill the requirements of efficient adsorption and thermal desorption for higher molecular weight compounds. The temperature stability of Tenax GC is relatively high (up to 400°C), and it is widely used for headspace sample analysis.[19-21,37,38]

Glass wool (Pyrex brand) was obtained from Corning Glass Works, Corning, New York.

All of the above were washed with distilled ether and conditioned at 280°C in a stream of dry helium for 12 hours.

2-chloropropane (Eastman-Kodak, Rochester, New York) and n-pentane (J. T. Baker, Phillipsburg, New Jersey) were distilled from phosphorous pentoxide (J. T. Baker, Phillipsburg, New Jersey). Diethyl ether anhydrous (Mallinckrodt, St. Louis, Missouri) was distilled from lithium aluminum hydride (Alfa Products, Danvers, Massachusetts) and water was distilled from potassium permanganate (Fisher Scientific, Pittsburgh, Pennsylvania).

B. Serum Samples

A total of thirty-six serum samples from twelve male volunteers who had been exposed to either "England" or "Rhinovirus" were supplied by Dr. H. J. Schneider (NASA, Houston, Texas, U.S.A.). These samples form part of a study of immune response to respiratory virus infection conducted by Dr. R. Couch, Dr. V. Knight and Dr. S. Criswell of the Influenza Research Center (Houston, Texas). Serum samples (200-800 µl) were stored in glass vials at -20°C prior to analysis.

The serum samples fall into two categories. The first (Nos. 1 → 6) were obtained from volunteers who developed clinical symptoms of influenza after virus infection. The second group (Nos. 7 → 12) were obtained from volunteers who did not develop clinical symptoms after virus infection. Each volunteer provided three serum samples corresponding to a baseline sample taken prior to virus infection (Group I), a serum sample taken one day after exposure to the virus (Group II) and a final

sample taken 14-21 days after virus infection (Group III). Group (I) represents normal serum samples, Group (II) virus infected serum samples (irrespective of whether clinical symptoms developed after exposure) and Group (III) virus infected serum samples from which all volunteers had recovered clinically from the infection.

C. Transevaporator Sampling Method

1. Sampling Apparatus

The transevaporator used is shown in Figure 1 and is essentially of the same design described by Zlatkis and Kim.[58] Significant differences are that the dimension of the Porasil E tube is changed to smaller size (70 mm x 2 mm I.D.) and that a water condenser is included.

The micro elution column was a glass tube containing 0.3 ml of Porasil E with a plug of glass wool at both ends. Prior to packing, Porasil E was conditioned again at 300°C by purging with nitrogen at 20 ml/min for 3 min. The glass bead trap (3.4 ml of glass beads) and Tenax trap (1.8 ml of Tenax GC) serve as a concentrating column, in a glass tube (110 mm x 10 mm O.D., 8 mm I.D.), tightly packed with glass wool. The elution apparatus consists basically of a custom made "transevaporator" and a conical tip centrifuge tube with a glass stopper (Ŝ 10/18) as shown in Figure 1.

2. Sampling Procedure

Figure 2 is a schematic diagram of a combined headspace and extraction sampling procedure. The sampling procedure is performed in

FIGURE 1. MODIFIED "TRANSEVAPORATOR" SAMPLING APPARATUS

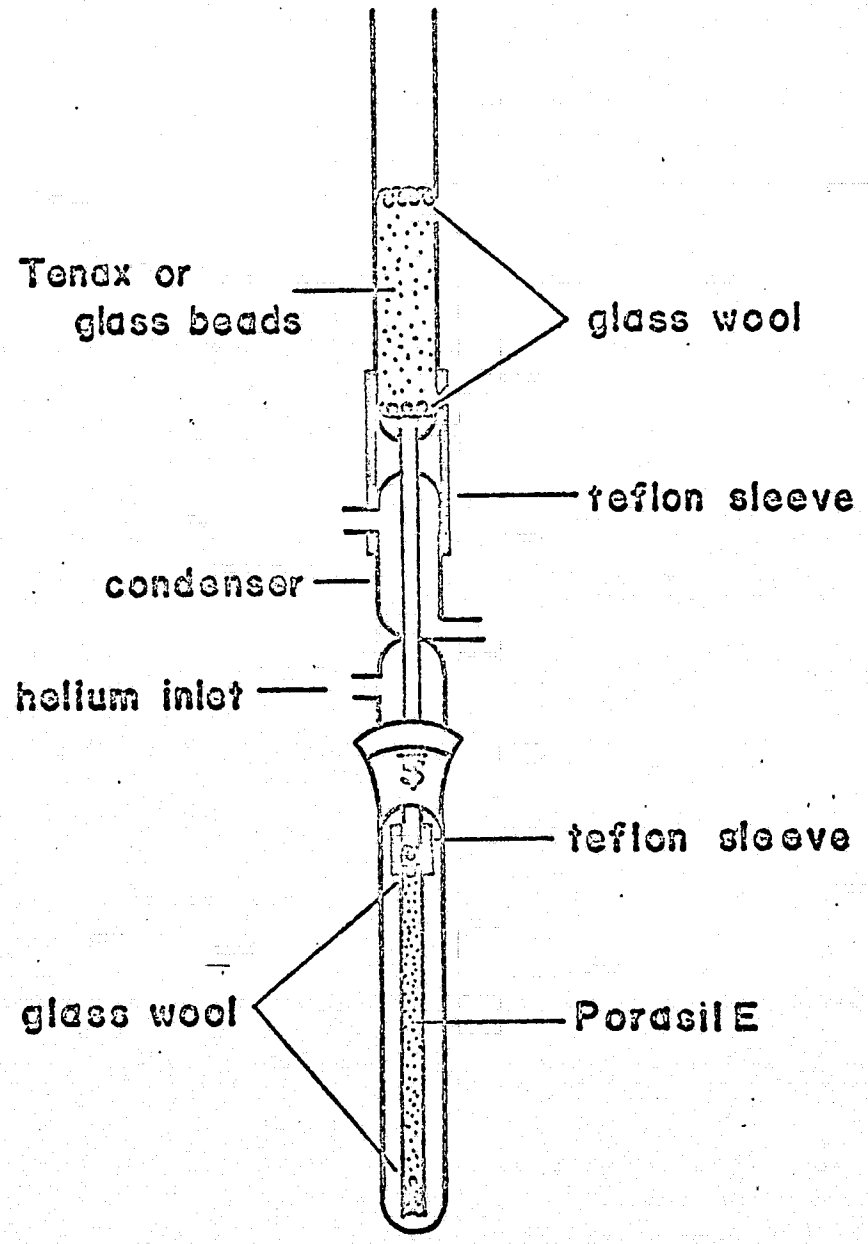
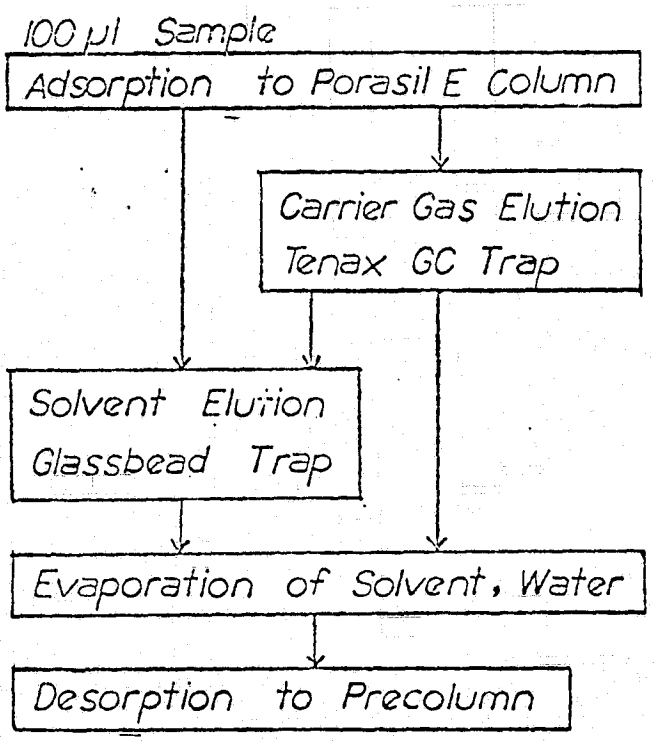

Tenax or
glass beads
glass wool
teflon sleeve
condenser
helium inlet
teflon sleeve
glass wool
Porasil E

FIGURE 2. SCHEMATIC DIAGRAM OF A COMBINED HEADSPACE AND EXTRACTION SAMPLING PROCEDURE


100 µl Sample
Adsorption to Porasil E Column
Carrier Gas Elution
Tenax GC Trap
Solvent Elution
Glassbead Trap
Evaporation of Solvent, Water
Desorption to Precolumn

two steps, the first being collection of low molecular weight volatiles on a Tenax tube, the second being the solvent extraction of volatiles coupled with collection on glass beads.

In the first step the sample (25 to 200 µl) is introduced to the inner tube which contains 0.3 ml of Porasil E, a strongly hydrophilic porous silica adsorbent. The upper tube contains 1.8 ml of Tenax, a 2,6-diphenyl-p-phenylene oxide porous polymer, which is an excellent trapping material for headspace analysis. A condenser is included in the apparatus to prevent excess water vapor from reaching the Tenax trap. A stream of helium gas (16 ml/min) is passed through the apparatus for 5 to 10 minutes to transfer the volatile compounds from the porasil tube to the Tenax tube which is then removed from the apparatus. Helium (16 ml/min) is passed through the tube for about 3 minutes to remove traces of water which would block the precolumn during desorption. There is no apparent change in the resultant profile due to this procedure. The Tenax tube is removed from the apparatus and placed in a desorber at 280°C for 10 minutes. A stream of helium (7 ml/min) desorbs the volatiles into a precolumn, cooled in liquid nitrogen, which is then connected to an analytical capillary column for chromatography.

In the second step the apparatus is converted to the form described by Zlatkis and Kim[58] by removing the condenser. The Tenax tube is exchanged for a tube of identical dimensions containing glass beads. 0.8 ml of 2-chloropropane is introduced into the tube and is swept through the porasil tube into the upper tube by applying gas pressure. Soluble organics are extracted and transferred to the glass beads while most of

the water and high molecular weight material is retained by the Porasil E. The Porasil E tube is removed, the apparatus is reassembled and the outer tube is placed in an air bath maintained at 50°C to heat incoming helium which is passed through the apparatus at a rate of 16 ml/min for 10 minutes to remove excess 2-chloropropane and traces of water. As in the case with the Tenax tube there is no apparent change in the resultant profile due to this procedure. The organics are transferred from the glass bead trap to the capillary column by the same technique as is used for the Tenax tube.

3. Transference of Sample from Trapping Tube to Analytical Column

The trapping tube was placed in a desorption chamber maintained at 280°C, connected to a stainless steel precolumn (30 cm x 1 mm I.D., coated with SF-96) and cooled in liquid nitrogen. The volatiles were transferred to the precolumn by passing helium through the system at 7 ml/min for 10 minutes. The precolumn was then connected to the analytical capillary column, the first 30 cm of which was cooled in liquid nitrogen. The sample was transferred from the precolumn to the analytical column by heating the former to 180°C with an air heating gun for 1 minute while passing helium through the system at a rate of 1.5 ml/min.

D. Capillary Column Gas Chromatography

1. Analytical Capillary Column

In this study, all analyses were carried out on stainless steel capillary columns (100 m x 0.25 mm, I.D.) which were purchased from

Handy and Harman Tube Co., Norristown, Pennsylvania. Newly acquired columns were first checked for leaks and then modified at both ends by slipping short pieces of stainless steel tubing (40 mm x 1.0 mm I.D. and 1.6 mm O.D.) over the end, after silver soldering, the sleeve was fitted with 1/16 in. ferrule and nut.

Cleaning of the new columns was done by washing with several 25 ml portions of acetone, n-pentane , chloroform, alcohol, water and acetone at 500 psi. After drying the column at 150°C for one hour, the quality of the empty column was tested by injecting a series of test substances (cyclohexane, benzene, diethyl ketone and ethanol) individually at 80°C. All test substances should be free of significant tailing. To improve the quality of empty columns which show severe tailing, the column was filled with etching solution (5 g ferric chloride, 15 ml conc. HCl and 60 ml distilled water) for 2 hrs and washed with distilled water several times, followed by solvent washing.

The liquid phase used in this study was 5% Witconol LA-23 (Witco, Houston, Texas), an alkyl aryl polyether alcohol, in CH_2Cl_2 solution (W/W). A post column of identical I.D. is attached at one end of the analytical column and the other end is immersed in a beaker filled with water. The flow rate throughout the coating process was controlled at one bubble per second. After drying overnight at room temperature by passing nitrogen carrier gas the coated column was conditioned in the gas chromatograph at a low flow rate and temperature programmed at a rate of 0.5°C/min from 40°C to 160°C.

The efficiency of a coated column was checked with a decalin mixture (decalin:diamyl ether = 1:1, V/V) at 120°C, isothermally. The resolving power was also tested with an aromatic mixture (cyclohexane, benzene, ethylbenzene, meta- and para-xylene) at 80°C, isothermally. Most of the columns which were coated with this method gave 250,000 theoretical plates and provided base line separation between meta- and para-xylene.

2. Gas Chromatographic Conditions

A Hewlett-Packard 5830 A gas chromatograph (Hewlett-Packard, Avondale, Pennsylvania) equipped with flame ionization detectors and connected to a Hewlett-Packard 18850 gas chromatography terminal was used. The serum volatiles were separated on a stainless steel (100 m x 0.25 mm I.D.) capillary column coated with Witconol LA-23. With a helium carrier gas flow rate of 1.5 ml min^{-1}, the column was held isothermally at 50°C for ten minutes and then programmed at 1.5°C min^{-1} to 160°C and maintained at this temperature for eighty minutes.

3. Gas Chromatography-Mass Spectrometry (GC-MS)

Gas chromatography-mass spectrometry (GC-MS) was performed on an LKB 9000 instrument (LKB-Prodkter AB, S-161, 25 Bromma 1, Sweden) with a single jet separator and a Perkin-Elmer 900 gas chromatograph. The ionization voltage was 70 eV and the scan time for a mass range of 20 to 300 atomic mass units was 4 seconds. The temperature of the separator and ionic source were set 220°C and 250°C respectively. Analytical procedures and chromatographic conditions were the same as described in the previous section. When possible identification was confirmed by

comparison with standard compounds available in the laboratory, otherwise manual interpretation and comparison to library spectra[75,76] was used.

E. Data Handling and Computer Interpretation

The operations which constitute the interpretative procedure are summarized in Figure 3. Visual inspection was used to identify those peaks common to all chromatograms. The peaks were normalized and this formed the data set for calculation of system variations and variation between individuals within each group. A second data set selected on the basis of the magnitude of their difference between groups was used to calculate all possible peak ratio combinations and diagnostic ratios selected. These ratios were autoscaled and used as input for K-nearest neighbor classification

1. Autoscaling[32,67]

In autoscaling, each dimension was adjusted to a mean of zero with unit standard deviation. Hence:

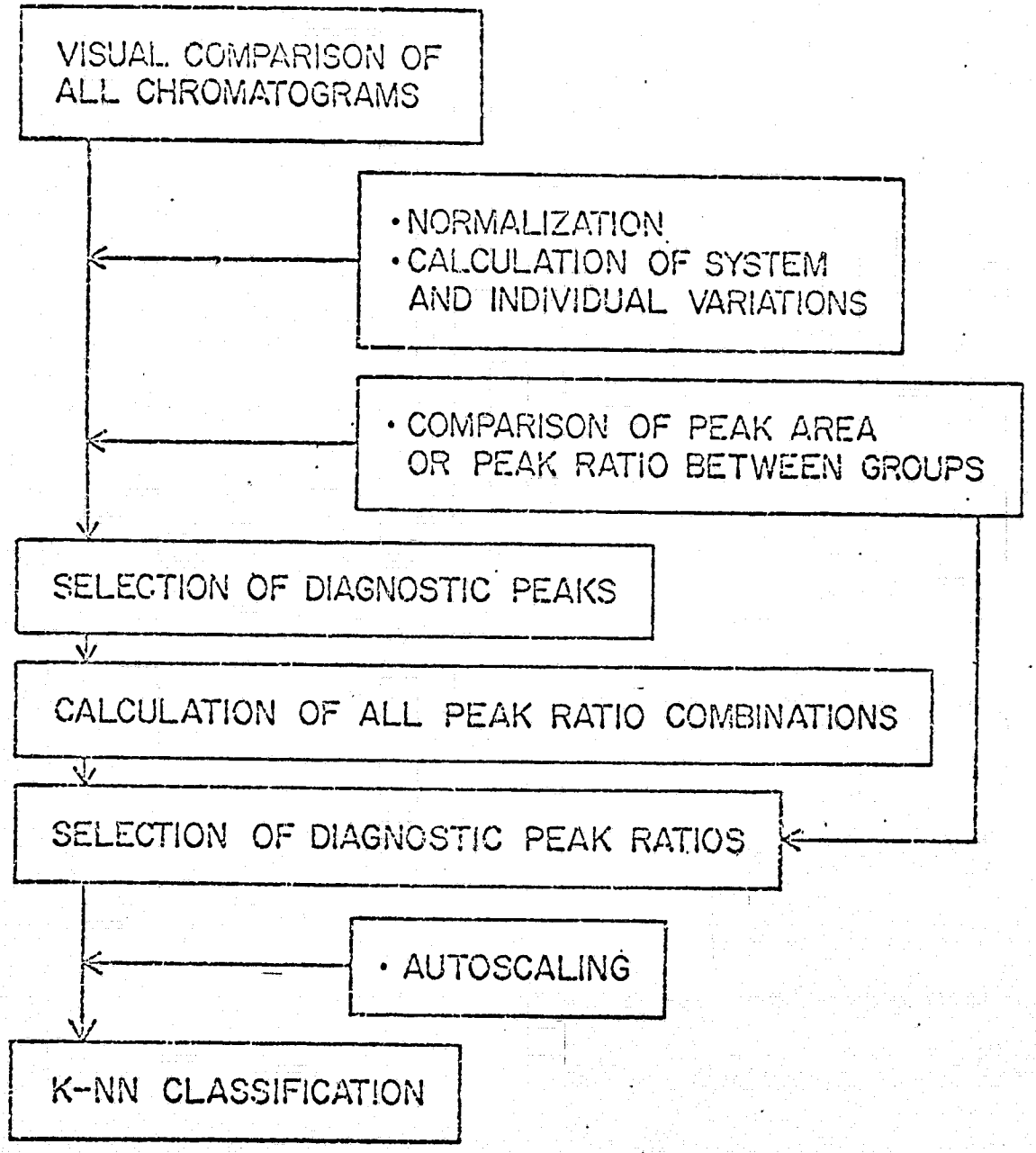
$$X'_{im} = \frac{X_{im} - \bar{X}_m}{\sigma_m}$$

where $\bar{X}_m$ is the mean and σ_m is the standard deviation of the value X_{im} for the ith feature. X'_{im} is the new autoscaled value. This preprocessing minimizes the effect of differing data magnitude. It was known that autoscaled data gives high rates of correct classification in K-NN method.

2. K-Nearest-Neighbor (KNN) Classification[32,63-67]

This pattern classification technique is based upon the assumption

FIGURE 3. SCHEMATIC DIAGRAM OF DATA PROCESSING OF CHROMATOGRAPHIC INFORMATION


VISUAL COMPARISON OF ALL CHROMATOGRAMS
• NORMALIZATION
• CALCULATION OF SYSTEM AND INDIVIDUAL VARIATIONS
• COMPARISON OF PEAK AREA OR PEAK RATIO BETWEEN GROUPS
SELECTION OF DIAGNOSTIC PEAKS
CALCULATION OF ALL PEAK RATIO COMBINATIONS
SELECTION OF DIAGNOSTIC PEAK RATIOS
• AUTOSCALING
K-NN CLASSIFICATION

that "the closer together" patterns are, the more closely related are the sets of data. Closeness was defined as the distance between an unknown data point and points of a known group (training set). The distance between point 1 and m is defined in this case as a Euclidian distance:

$$d_{1m} = [\sum_{k=1}^{n} (X_{1k} - X_{mk})^2]^{1/2}$$

where k is the number of dimensions in factor space. In the case of the single (k = 1) nearest neighbor (1-NN), the unknown point is classified as the group which gives the smallest distance from the points used for the training set.

3. Computer Programs for Profile Analysis

All programs were written in Fortran IV and run on the University of Houston computing center's Honeywell 66/60. These computer programs consisted of a short main program which inputs a set of data, initializes a number of parameters, and calls the several subroutine subprograms; percent normalization, calculation of relative standard deviation, writing, calculation of all possible combinations of two peak ratios, autoscaling, sorting in ascending order and classification by K-nearest neighbor methods.

The array A contains the raw integrated peak area. It is dimensioned for up to 36 chromatograms (N) of up to 37 selected peak areas (M) per chromatogram. The following is an example of an actual computer program used in this study.

```
C
C ***************************************************************
C ***************************************************************
C  PROFILE ANALYSIS OF VIRUS INFECTED SERUM
C  DATA, NORMALIZED DATA,PEAK RATIO, CALCULATION OF
C  MEAN, SD., PT. COEF. VAR., AND AUTOSCALING
C ***************************************************************
C ***************************************************************
C
C                    AUG. 10, 1978  K.Y.LEE
C
C
C PROFILE ANALYSIS OF BIOLOGICAL FLUID
C
C  N = NUMBER OF COLUMNS (SAMPLE NUMBERS)
C  M = NUMBER OF ROWS (SELECTED PEAKS OR PEAK RATIO)
C  A = DATA POINTS ARRAY; A(M,N)
C  B = PERCENT NORMALIZED DATA POINTS; B(M,N)
C  R = PEAK RATIO; R(M,N)
C  ASLB = AUTOSCALED POINTS ; ASLB(M,N)
C
      DOUBLE PRECISION B,ASLB,MEANB,SDB,PTCV,R
      DIMENSION A(37,36),B(37,36),ASLB(37,36),MEANB(37),
     1          SDB(37),PTCV(37),R(7,36)
C
      READ(5,100) N,M
C   INITIALIZE DATA POINTS
      DO 10 I = 1,M
      DO 10 J = 1, N
      A(I,J)=0.0
      B(I,J)=0.0
      R(I,J)=0.0
      ASLB(I,J)=0.0
      MEANB(I)=0.0
      SDB(I)=0.0
      PTCV(I)=0.0
   10 CONTINUE
C
      WRITE (6,118)
      WRITE (6,119)
      DO 15 I = 1,M
      READ (5,101) (A(I,J), J=1,N)
      WRITE (6,121) I, (A(I,K), K=1,N)
   15 CONTINUE
C
C  NORMALIZATION(PERCENT) OF DATA
      CALL NORM (A,N,M,B)
      CALL WRITE (B,N,M)
C  CALCULATION OF REPRODUCIBILITY
      WRITE (6,350)
      L=1
      NN = 12
```

ORIGINAL PAGE IS
OF POOR QUALITY

16.421

```
      KL = 3
      CALL CALC (B,L,N,NN,KL,M,MEANB,SDB,PTCV )
      WRITE (6,340)
      L = 2
      CALL CALC (B,L,N,NN,KL,M,MEANB,SDB,PTCV )
      WRITE (6,350)
      L = 3
      CALL CALC (B,L,N,NN,KL,M,MEANB,SDB,PTCV )
  100 FORMAT(2I10)
  101 FORMAT ( 3(9F8.0,/),9F8.0 )
  118 FORMAT (1H1,//,"******** DATA POINTS **************
     1**************************************",//)
  119 FORMAT (//, "PEAK",2X,"1-1",4X,"1-2",4X,"1-3",4X,
     1          "2-1",4X,"2-2",4X,"2-3",4X,"3-1",4X,"3-2" ,
     1          4X,"3-3",4X,"4-1",4X,"4-2",4X,"4-3",4X,
     1          "5-1",4X,"5-2",4X,"5-3",4X,"6-1",4X,"6-2",
     1          4X,"6-3",4X,//,6X,"7-1",4X,"7-2",4X,"7-3",
     1          4X,"8-1",4X,"8-2",4X,"8-3",4X,"9-1",4X,
     1          "9-2",4X,"9-3",4X,"10-1",3X,"10-2",3X,
     1          "10-3",3X,
     1          "11-1",3X,"11-2",3X,"11-3",3X,"12-1",3X,
     1          "12-2",3X,"12-3", /// )
  121 FORMAT (I3, 18F7.0,/,3X, 18F7.0, // )
  330 FORMAT(1H1,//,"**BEFORE INFECTION 12 SAMPLES**",//)
  340 FORMAT(1H1,//,"**AFTER  INFECTION 12 SAMPLES**",//)
  350 FORMAT(1H1,//,"**AFTER  CURE  ,   12 SAMPLES**",//)
      STOP
      END
```

ORIGINAL PAGE IS
OF POOR QUALITY

```
C
C ***************************************************
C ***************************************************
C POOLED VIRUS SAMPLE, REPRODUCIBILITY
C NORMALIZATION, 6 SAMPLES
C  PEAK RATIO, AND MEAN , SD, PT, COEF. VAR.
C ***************************************************
C ***************************************************
C
C                    AUG. 10, 1978  K.Y.LEE
C
C PROFILE ANALYSIS OF BIOLOGICAL FLUID
C
C  N = NUMBER OF COLUMNS (SAMPLE NUMBERS)
C  M = NUMBER OF ROWS (SELECTED PEAKS)
C  A = DATA POINTS ARRAY; A(M,N)
C  B = PERCENT NORMALIZED DATA POINTS; B(M,N)
C  ASLB = AUTOSCALED POINTS ; ASLB(M,N)
C  R = PEAK RATIO
C
      DOUBLE PRECISION B,ASLB,MEANB,SDB,PTCV,R
      DIMENSION A(37,7),B(37,7),ASLB(37,7),MEANB(37),
     1          SDB(37),PTCV(37),R(2,7)
C
      READ(5,100) N,M
C   INITIALIZE DATA POINTS
      DO 10 I = 1,M
      DO 10 J = 1, N
      A(I,J)=0.0
      B(I,J)=0.0
      R(I,J) = 0.0
      ASLB(I,J)=0.0
      MEANB(I)=0.0
      SDB(I)=0.0
      PTCV(I)=0.0
   10 CONTINUE
C
      WRITE (6,120)
      DO 20 I = 1, M
      READ (5, 101)  (A(I,J), J=1,N)
      WRITE (6,121) I,(A(I,K),K=1,N)
   20 CONTINUE
C
C  NORMALIZATION(PERCENT) OF DATA
            CALL NORM ( A,N,M,B )
C  CALCULATION OF MEAN, S.D., PERCENT COEFF. VAR.
      L = 1
      NN = 6
      KL=1
      CALL CALC (B,L,N,NN,KL,M,MEANB,SDB,PTCV )
C CALCULATION OF PEAK RATIO OF POOLED SWRUM
      WRITE (6,122)
```

16.426

```
      DO 23 I = 1,2
      DO 25 J = 1, N
      IF ( J .EQ. 6) GO TO 25
      IF ( I .EQ. 1 ) GO TO 28
      IF ( I .EQ. 2 ) GO TO 29
   28 R(I,J)=B(16,J) / B(25,J)
      GO TO 25
   29 R(I,J) = B(29,J) / B(34,J)
   25 CONTINUE
      WRITE (6,123)  I, (R(I,L), L=1,N)
   23 CONTINUE
C  CALCULATION OF MEAN, S.D., PERCENT COEFF. VAR. OF RATIO
      M=2
      CALL CALCI(R,L,N,NN,KL,M,MEANB,SDB,PTCV )
C  AUTOSCALING OF PEAK RATIO
            CALL AUTS (R,N,M,MEANB,SDB,ASLB )
C  CALCULATION OF MEAN,SD, PT.COEF.VAR. OF AUTOSCALED POINT
      CALL CALCI (ASLB,L,N,NN,KL,M,MEANB,SDB,PTCV)
  100 FORMAT(2I10)
  101 FORMAT ( 7F10.0 )
  120 FORMAT ('1',6X,"***DATA POINTS***",/,11X,"PEAK NO",
     1         5X,"SAMP 1",8X,"SAMP 2",8X,"SAMP 3",8X,
     1         "SAMP 4",8X,"SAMP 5",8X,"SAMP 6",8X,"SAMP 7",
     1         // )
  121 FORMAT (/,10X,I5,7F14.0)
  122 FORMAT (1H1,///,6X,"**PEAK RATIO OF  1/4, 5/7 ****
     1 **************************",//,11X,"PEAK NO",
     1         "SAMP 4",8X,"SAMP 5",8X,"SAMP 6",8X,"SAMP 7",
     1         5X,"SAMP 1",8X,"SAMP 2",8X,"SAMP 3",8X,
     1         // )
  123 FORMAT ( 10X,I5, 7F14.2, // )
      STOP
      END
```

```
C
C ***************************************************************
C ***************************************************************
C PROFILE ANALYSIS OF BEFORE,AFTER AND CURE VIRUS SAMPLE
C SELECTED 7 PEAKS(NORMALIZED)    P16=1,   P19=2,
C          P20=3,   P25=4,   P29=5,   P31=6,   P34=7,
C          RESPECTIVELY AND  CALCULATION OF PEAK RATIO
C REPRODUCIBILITY OF PEAK RATIO WITHIN GROP
C ***************************************************************
C ***************************************************************
C
C                  AUG. 10, 1978  K.Y.LEE
C
C PROFILE ANALYSIS OF BIOLOGICAL FLUID
C
C  N = NUMBER OF COLUMNS (SAMPLE NUMBERS)
C  M = NUMBER OF ROWS (SELECTED PEAKS)
C  B = PERCENT NORMALIZED DATA POINTS; B(M,N)
C  ASLB = AUTOSCALED POINTS; ASLB(M,N)
C  R = PEAK RATIO ; R(M,N)
C
      DOUBLE PRECISION B,ASLB,R,MEANB,SDB,PTCV
      DIMENSION B(7,36),ASLB(7,36),R(7,36),
     1          MEANB(7),SDB(7),PTCV(7)
C
      READ(5,100) N,M
C   INITIALIZE DATA POINTS
      DO 10 I = 1,M
      DO 10 J = 1, N
      B(I,J)=0.0
      R(I,J)=0.0
      ASLB(I,J)=0.0
      MEANB(I)=0.0
      SDB(I)=0.0
      PTCV(I)=0.0
   10 CONTINUE
C
      WRITE (6,120)
      DO 15 I = 1,M
      READ (5,101) (B(I,J), J=1,N)
      WRITE (6,121) I, (B(I,K), K=1,N)
   15 CONTINUE
C
      WRITE (6,220)
      CALL WRITE (B,N,M)
C  CALCULATION OF PEAK RATIO(NORMALIZED)
      CALL RATO(B,N,M,R)
  100 FORMAT(2I10)
  101 FORMAT ( 3(10F8.2,/), 6F8.2 )
  120 FORMAT (1H1,//,8X,"**SELECTED 7 PEAKS, NORMALIZED
     1 **********************************",///,
     1            "PEAK",2X,"1-1",4X,"1-2",4X,"1-3",4X,
```

16.431

```
     1          "2-1",4X,"2-2",4X,"2-3",4X,"3-1",4X,"3-2" ,
     1          4X,"3-3",4X,"4-1",4X,"4-2",4X,"4-3",4X,
     1          "5-1",4X,"5-2",4X,"5-3",4X,"6-1",4X,"6-2",
     1          4X,"6-3",4X,//,6X,"7-1",4X,"7-2",4X,"7-3",
     1          4X,"8-1",4X,"8-2",4X,"8-3",4X,"9-1",4X,
     1          "9-2",4X,"9-3",4X,"10-1",3X,"10-2",3X,
     1          "10-3",3X,
     1          "11-1",3X,"11-2",3X,"11-3",3X,"12-1",3X,
     1          "12-2",3X,"12-3", /// )
  121 FORMAT (I3, 18F7.2,/,3X, 18F7.2, ///)
  220 FORMAT (1H1,//,"*****NORMALIZED DATA, 7 SAMPLES ****
     1**********************************************",///)
      STOP
      END
```

ORIGINAL PAGE IS
OF POOR QUALITY

```
C
C ***********************************************************
C ***********************************************************
C PROFILE ANALYSIS
C  AUTOSCALING OF SELECTED 2 PEAK RATIO, 1/4 AND 5/7
C ***********************************************************
C ***********************************************************
C
C                    AUG. 10, 1978  K.Y.LEE
C
C
C PROFILE ANALYSIS OF BIOLOGICAL FLUID
C
C  N = NUMBER OF COLUMNS (SAMPLE NUMBERS)
C  M = NUMBER OF ROWS (SELECTED PEAKS OR PEAK RATIO)
C  R = PEAK RATIO DATA; R(M,N)
C  ASLB = AUTOSCALED POINTS ; ASLB(M,N)
C
      DOUBLE PRECISION R,ASLB,MEANB,SDB,PTCV
      DIMENSION R(2,36),ASLB(2,36),MEANB(2),SDB(2),PTCV(2)
C
      READ(5,100) N,M
C   INITIALIZE DATA POINTS
      DO 10 I = 1,M
      DO 10 J = 1, N
      R(I,J)=0.0
      ASLB(I,J)=0.0
      MEANB(I)=0.0
      SDB(I)=0.0
      PTCV(I)=0.0
   10 CONTINUE
C
      WRITE (6,120)
      DO 15 I = 1,M
      READ (5,101) (R(I,J), J=1,N)
      WRITE (6,121) I, (R(I,K), K=1,N)
   15 CONTINUE
C
C  WRITE  PEAK RATIO OF EACH GROUP
      WRITE (6,151)
      CALL WRITE (R,N,M)
C  CALCULATION OF MEAN, S.D., PERCENT COEFF. VAR.
      L = 1
      NN = 36
      KL = 1
      CALL CALC (R,L,N,NN,KL,M,MEANB,SDB,PTCV )
C  AUTOSCALING OF PEAK RATIO; MEAN=0.0, S.D.=1.0
      CALL AUTS (R,N,M,MEANB,SDB,ASLB )
      WRITE (6,126)
      CALL WRITE (ASLB,N,M )
C
  100 FORMAT(2I10)
```

16.429

```
  101 FORMAT ( 3(10F8.2,/), 6F8.2 )
  120 FORMAT (1H1,//,8X,"** SELECTED 2 PEAK RATIO**",
     1          "1/4 AND 5/7, **********************",//,
     1              "PK RTO" ,"1-1",4X,"1-2",4X,"1-3",4X,
     1          "2-1",4X,"2-2",4X,"2-3",4X,"3-1",4X,"3-2" ,
     1          4X,"3-3",4X,"4-1",4X,"4-2",4X,"4-3",4X,
     1          "5-1",4X,"5-2",4X,"5-3",4X,"6-1",4X,"6-2",
     1          4X,"6-3",4X,//,6X,"7-1",4X,"7-2",4X,"7-3",
     1          4X,"8-1",4X,"8-2",4X,"8-3",4X,"9-1",4X,
     1          "9-2",4X,"9-3",4X,"10-1",3X,"10-2",3X,
     1          "10-3",3X,
     1          "11-1",3X,"11-2",3X,"11-3",3X,"12-1",3X,
     1          "12-2",3X,"12-3", /// )
  121 FORMAT (I3, 18F7.2,/,3X, 16F7.2, //////)
  126 FORMAT ('1',6X,"**AUTOSCALING OF PEAK RATIO******
     1 **********************************************",//)
  151 FORMAT (1H1,//, "** SELECTED 2 PEAK RATIO *********
     1**********************************************",//)
      STOP
      END
```

```
C
C ***************************************************
C ***************************************************
C  K-NN PATTERN RECOGNITION
C  NORMAL OR VIRUS INFECTED SERUM
C ***************************************************
C ***************************************************
C
C                    AUG. 10, 1978  K.Y.LEE
C
C PROFILE ANALYSIS OF BIOLOGICAL FLUID ,K-NN CLASSIFICATION
C
C  M = NO. OF UNKNOWN SAMPLES(ROW): M(I)
C  L = NO. OF TRAINING SET(ROW) ; L(J)
C  N = NO. OF FACTOR SPACE (COLUMN) ; N(K), K=2
C  X = UNKNOWN SAMPLE SET; X(I,K)
C  Y = KNOWN TRAINING SET ; Y(J,K)
C  D = EUCLIDEAN DISTANCE IN N(K) SPACE BETWEEN POINTI,J
C
      DOUBLE PRECISION X,Y,D
      DIMENSION X(26,2), Y(10,2), D(26,10),JJ(10)
C
      READ (5,100) L,M,N
C   INITIALIZE DATA POINTS
      DO 10 I = 1,M
      DO 11 J = 1,L
      DO 12 K = 1,N
      X(I,K)=0.0
      Y(J,K)=0.0
      D(I,J)=0.0
   12 CONTINUE
   11 CONTINUE
   10 CONTINUE
C
      WRITE (6,110)
      DO 15 I = 1,M
      READ (5,105)   (X(I,K), K=1,N)
      WRITE (6,112) I, (X(I,K), K=1,N)
   15 CONTINUE
      WRITE (6,113)
      DO 17 J = 1,L
      READ (5,106) (Y(J,K), K=1,N)
      WRITE (6,112) J, (Y(J,K), K=1,N)
   17 CONTINUE
C
C  CALCULATION OF EUCLIDEAN DISTANCE IN  2 FACTOR SPACE
      WRITE (6,220)
      DO 20 I = 1,M
      DO 30 J = 1,L
      JJ(J)=J
      D(I,J)=SQRT((X(I,1)-Y(J,1))**2+ (X(I,2)-Y(J,2))**2)
   30 CONTINUE
```

17.139

```
      WRITE (6,114) I, (D(I,K), K=1,L )
   20 CONTINUE
      DO 31 I =1,M
      WRITE (6,221) I
      DO 32 J=1,L
   32 WRITE (6,222) JJ(J), D(I,J)
   31 CONTINUE
C  SORTING AND CLASSIFICATION BY K-NN METHOD
      CALL SORT (JJ,D,M,L )
  100 FORMAT (3I12 )
  105 FORMAT ( 2F10.2 )
  106 FORMAT ( 2F8.2 )
  110 FORMAT (1H1,//,"***UNKNOWN SAMPLE SET ************
     1 ****************************************",//,
     1        "SAMPLE NO",5X,"X(I,1)",4X,"X(I,2)",//)
  112 FORMAT ( I10, 2F10.2,// )
  113 FORMAT (1H1,//,"***TRAINING SET FOR K-NN **********
     1   BEFORE AND AFTER INFECTION ******************",//,
     1        "TRAINING ST",4X,"Y(J,1)",4X,"Y(J,2)",//)
  114 FORMAT (I7,5F9.3,7X,5F9.3, //)
  220 FORMAT (1H1,//,"*** K-NN CLASSIFICATION **********",
     1 /,"******EUCLIDEAN DISTANCE D(I,J) IN 2 FACTOR **
     1 ****************************************",//,17X,
     1 "**TRAINING SET(NORMAL)**",20X,"TRAINING SET(INFECT)
     1 **",/,"SAMP NO",7X,"B1",7X,"B4",6X,"B7",6X,"B10",
     1 6X,"B12",15X,"A1",7X,"A2",7X,"A3",6X,"A7",6X,"A8",
     1  // )
  221 FORMAT (///,"***SAMP=",I3," *****", / )
  222 FORMAT (10X, 10(I4, F8.3 ), // )
      STOP
      END
```

16.422 UBROUTINE PROGRAM FOR WRITING

```
C       SUBROUTINE PROGRAM FOR WRITING
        SUBROUTINE WRITE (R,N,M )
C
        DOUBLE PRECISION R
        DIMENSION R(2,36)
C
C BEFORE INFECTION,AFTER INFECTION AND AFTER CURE
C SAMPLE NO FROM 1 TO 12,
        WRITE (6,130)
        DO 35 I = 1,M
   35 WRITE (6,131) I,(R(I,L), L=1,36,3 )
        WRITE (6,140)
        DO 40 I = 1,M
   40 WRITE (6,131) I, (R(I,L), L=2,36,3 )
        WRITE (6,150)
        DO 50 I = 1,M
   50 WRITE (6,131) I, ( R(I,L), L=3, 36, 3 )
C
C SAMPLE NO FROM 1 TO 6 ,ILL VIRUS SAMPLE
        WRITE (6,160)
        DO 60 I = 1,M
   60 WRITE (6,161) I, ( R(I,L), L=1,9 )
        WRITE (6,170)
        DO 70 I = 1,M
   70 WRITE (6,161) I, ( R(I,L), L=10,18 )
C
C SAMPLE NO FROM 7 TO 12, NOT ILL VIRUS SAMPLE
        WRITE (6,260)
        DO 62 I = 1,M
   62 WRITE (6,161) I, ( R(I,L), L=19,27)
        WRITE (6,270)
        DO 72 I = 1,M
   72 WRITE (6,161) I, ( R(I,L), L=28,36)
C
  130 FORMAT (  ////,8X,"**BEFORE INFECTION, 12 SAMPLES
     1          **",//,"PEAK",5X,"1-1",6X,"2-1",6X,"3-1",
     1          6X,"4-1",6X,"5-1", 6X,"6-1", 6X,"7-1", 6X,
     1          "8-1", 6X,"9-1", 5X,"10-1", 5X,"11-1",
     1           5X,"12-1",// )
  131 FORMAT (I3, 12F9.2 ,// )
  140 FORMAT (  ////,8X,"**AFTER  INFECTION, 12 SAMPLES
     1          **",//,"PEAK",5X,"1-2",6X,"2-2",6X,"3-2",
     1          6X,"4-2",6X,"5-2", 6X,"6-2", 6X,"7-2", 6X,
     1          "8-2", 6X,"9-2", 5X,"10-2", 5X,"11-2",
     1           5X,"12-2",// )
  150 FORMAT (  ////,8X,"**AFTER  CURE      , 12 SAMPLES
     1          **",//,"PEAK",5X,"1-3",6X,"2-3",6X,"3-3",
     1          6X,"4-3",6X,"5-3", 6X,"6-3", 6X,"7-3", 6X,
     1          "8-3", 6X,"9-3", 5X,"10-3", 5X,"11-3",
     1           5X,"12-3",// )
  160 FORMAT (1H1,//,8X," ****** ILL , 6 SAMPLES ******
     1**********************************************",//,
```

```
16.422   UBROUTINE PROGRAM FOR WRITING

     1          "PEAK",5X,"1-1",6X,"1-2",6X,"1-3",16X ,
     1          "2-1",6X,"2-2", 6X,"2-3",16X,"3-1", 6X,
     1          "3-2", 6X,"3-3",//)
  161 FORMAT ( I3, 3F9.2,10X,3F9.2,10X,3F9.2 ,//)
  170 FORMAT (  ////,8X," ** *** ILL , 6 SAMPLES **",//,
     1          "PEAK",5X,"4-1",6X,"4-2",6X,"4-3",16X,
     1          "5-1",6X,"5-2", 6X,"5-3",16X,"6-1", 6X,
     1          "6-2", 6X,"6-3",// )
  260 FORMAT(//////,8X,"**NOT ILL,6SAMPLES",//,"PEAK",
     1          5X,"7-1",6X."7-2",6X,"7-3",16X,"8-1",6X,
     1          "8-2",6X,"8-3",16X,"9-1",6X,"9-2",6X,"9-3",
     1           // )
  270 FORMAT(////,8X,"**NOT ILL 6 SAMPLES**",//,"PEAK",
     1           4X,"10-1",5X,"10-2",5X,"10-3",15X,"11-1",
     1          5X,"11-2",5X,"11-3",15X,"12-1",5X,"12-2",
     1          5X,"12-3",// )
      RETURN
      END
```

16.422 USBROUTINE PROGRAM FOR NORMALIZATION

```
C       SUBROUTINE PROGRAM FOR NORMALIZATION
        SU6ROUTINE NORM (A,N,M,B )
        DOUBLE PRECISION B,SUMA
        DIMENSION A(37,36),B(37,36),SUMA(36)
C
C    INITIALIZE DATA POINTS
        DO 10 J = 1, N
        SUMA(J)=0.0
     10 CONTINUE
C
        WRITE (6,120)
               DO 40 J = 1, N
               DO 50 I = 1, M
               SUMA(J) = SUMA(J) + A(I,J)
     50        CONTINUE
        WRITE (6,121) J, SUMA(J)
     40        CONTINUE
C
               WRITE (6,122)
               DO 60 I = 1, M
               DO 70 J = 1, N
               B(I,J) = ( A(I,J) * 100 ) / SUMA(J)
     70        CONTINUE
        WRITE (6,123) I, (B(I,K), K=1,N )
     60        CONTINUE
C
   120 FORMAT (1H1,///,"** TOTAL PEAK AREA ***************
      1*************************************",//,3X,
      1          "SAMP NO",5X,"TOTAL AREA",/// )
   121 FORMAT ( I10, 5X, F10.0, // )
   122 FORMAT (1H1,//,8X,"**PERCENT NORMALIZED DATA*******
      1*************************************",
      1          //, "PEAK",2X,"1-1",4X,"1-2",4X,"1-3",4X,
      1          "2-1",4X,"2-2",4X,"2-3",4X,"3-1",4X,"3-2" ,
      1          4X,"3-3",4X,"4-1",4X,"4-2",4X,"4-3",4X,
      1          "5-1",4X,"5-2",4X,"5-3",4X,"6-1",4X,"6-2",
      1          4X,"6-3",4X,//,6X,"7-1",4X,"7-2",4X,"7-3",
      1          4X,"8-1",4X,"8-2",4X,"8-3",4X,"9-1",4X,
      1          "9-2",4X,"9-3",4X,"10-1",3X,"10-2",3X,
      1          "10-3",3X,
      1          "11-1",3X,"11-2",3X,"11-3",3X,"12-1",3X,
      1          "12-2",3X,"12-3", /// )
   123 FORMAT (I3,18F7.2, /, 3X, 18F7.2, // )
               RETURN
               END
```

```
16.423   UBROUTINE PROGRAM FOR MEAN,S.D.,PERCENT COEF. VAR.

C       SUBROUTINE PROGRAM FOR MEAN,S.D.,PERCENT COEF. VAR.
        SUBROUTINE CALC (B,L,N,NN,KL,M,MEANB,SDB,PTCV )
        DOUBLE PRECISION SUMB,SDB,MEANB,PTCV,SSQB,BFTR ,B
        DIMENSION B(37,36),MEANB(37),SDB(37),PTCV(37),
       1          SUMB(37), SSQB(37),BFTR(37)
C
C    INITIALIZE DATA POINTS
        DO 10 I = 1,M
        SUMB(I)=0.0
        SSQB(I)=0.0
   10 CONTINUE
C
        WRITE (6,124)
               DO 80 I = 1,M
               DO 90 J = L,N,KL
               SUMB(I) = SUMB(I) + B(I,J)
               SSQB(I) = SSQB(I) + B(I,J) **2
   90          CONTINUE
               MEANB(I) = SUMB(I) / NN
               BFTR(I) = SQRT (NN*SSQB(I) - SUMB(I)**2 )
               SDB(I) = BFTR(I) / NN
               PTCV(I) = ( SDB(I) * 100 ) / MEANB(I)
               WRITE (6,125) I, MEANB(I), SDB(I), PTCV(I)
   80          CONTINUE
  124 FORMAT (///,5X,"**REPRODUCIBILITY OF SELECTED PEAKS
     1         **",//,11X,"PEAK NO",6X,"MEAN",10X,"S.D.",
     1         8X,"PT-C.V.",//)
  125 FORMAT ( 10X,I5, 3F14.2, // )
        RETURN
        END
```

```
16.424   UBROUTINE PRO  RAM FOR MEAN,S.D.  :RCENT COEF. VAR.

      C      SUBROUTINE PR    AM FOR MEAN,S.D    ERCENT COEF. VAR.
             SUBROUTINE CA    (R,L,N,NN,KL,M,!   VB;SDB,PTCV )
             DOUBLE PRECIS     SUMB,SDB,MEANB    :V,SSQB,BFTR ,R
             DIMENSION R(?     ),MEANB(2),SDB(   PTCV(2),SUMB(2),
            1          SSQB(   BFTR(2)
      C
      C    INITIALIZE DAT     INTS
             DO 10 I = 1,
             SUMB(I)=0.0
             SSQB(I)=0.0
        10 CONTINUE
      C
                      WRITE    ,124)
                      DO 80    = 1,M
                      DO 90    = L,N,KL
                      IF (     ,J) .GT. 18.00     0 TO 30
                      IF (     ,J)  .GT.  9.0:    GO TO 30
                      SUMB(    = SUMB(I) + R(
                      SSQB(    = SSQB(I) + R(      **2
                      GO TO
        30            NN =      1
        90            CONTI
                      MEANB(I) = SUMB(I) / NN
                      BFTR(I) = SQRT (NN*SSQB(I) - SUMB(I)**2 )
                      SDB(I) = BFTR(I) / NN
                      PTCV(I) = ( SDB(I) * 100 ) / MEANB(I)
                      WRITE (6,125) I, MEANB(I), SDB(I), PTCV(I)
        80            CONTINUE
       124 FORMAT (1H1,///,"***REPRODUCIBILITY OF SLED. PEAKS
          1 ********************************************
          1          **",//,11X,"PEAK NO",6X,"MEAN",10X,"S.D.",
          1          8X,"PT-C.V.",//)
       125 FORMAT (/,10X,I5,3F14.3,/ )
             RETURN
             END
```

16.423 UBROUTINE PROGRAM FOR CALCULATION OF PEAK RATIO

```
C       SUBROUTINE PROGRAM FOR CALCULATION OF PEAK RATIO
        SUBROUTINE RATO(B,N,M,R)
        DOUBLE PRECISION B,R
        DIMENSION B(7,36),R(7,36)
C
        DO 10 K = 1,M
        DO 20 I = 1,M
        DO 30 J = 1,N
        R(I,J) = B(I,J) / B(K,J)
   30 CONTINUE
   20 CONTINUE
C
C SAMPLE NO FROM 1 TO 12,  PEAK RATIO
C BEFORE INFECTION,AFTER INFECTION AND AFTER CURE
        WRITE (6,151) K
        WRITE (6,130)
        DO 35 I = 1,M
   35 WRITE (6,131) I,(R(I,L), L=1,36,3 )
        WRITE (6,151) K
        WRITE (6,140)
        DO 40 I = 1,M
   40 WRITE (6,131) I, (R(I,L), L=2,36,3 )
        WRITE (6,151) K
        WRITE (6,150)
        DO 50 I = 1,M
   50 WRITE (6,131) I, ( R(I,L), L=3, 36, 3 )
C
C SAMPLE NO FROM 1 TO 6 ,PEAK RATIO
C NOT ILL VIRUS SAMPLES
        WRITE (6,151) K
        WRITE (6,160)
        DO 60 I = 1,M
   60 WRITE (6,161) I, ( R(I,L), L=1,9 )
        WRITE (6,170)
        DO 70 I = 1,M
   70 WRITE (6,161) I, ( R(I,L), L=10,18 )
C
C SAMPLE NO FROM 7 TO 12,PEAK RATIO
C ILL VIRUS SAMPLES
        WRITE (6,151) K
        WRITE (6,260)
        DO 62 I = 1,M
   62 WRITE (6,161) I, ( R(I,L), L=19,27)
        WRITE (6,270)
        DO 72 I = 1,M
   72 WRITE (6,161) I, ( R(I,L), L=28,36)
   10 CONTINUE
C
  130 FORMAT (  ////,8X,"**BEFORE INFECTION, 12 SAMPLES
     1          **",//,"PEAK",5X,"1-1",6X,"2-1",6X,"3-1",
     1          6X,"4-1",6X,"5-1", 6X,"6-1", 6X,"7-1", 6X,
     1          "8-1", 6X,"9-1", 5X,"10-1", 5X,"11-1",
```

16.423 UBROUTINE PROGRAM FOR CALCULATION OF PEAK RATIO

```
     1          5X,"12-1",// )
 131 FORMAT (I3, 12F9.2 ,// )
 140 FORMAT (  ////,8X,"**AFTER  INFECTION, 12 SAMPLES
     1         **",//,"PEAK",5X,"1-2",6X,"2-2",6X,"3-2",
     1         6X,"4-2",6X,"5-2", 6X,"6-2", 6X,"7-2", 6X,
     1         "8-2", 6X,"9-2", 5X,"10-2", 5X,"11-2",
     1          5X,"12-2",// )
 150 FORMAT (  ////,8X,"**AFTER  CURE     , 12 SAMPLES
     1         **",//,"PEAK",5X,"1-3",6X,"2-3",6X,"3-3",
     1         6X,"4-3",6X,"5-3", 6X,"6-3", 6X,"7-3", 6X,
     1         "8-3", 6X,"9-3", 5X,"10-3", 5X,"11-3",
     1          5X,"12-3",// )
 151 FORMAT (1H1,//,2X,"K=",I2,3X,"** PEAK RATIO *******
     1**********************************",//)
 160 FORMAT ( // //,8X," ** NOT ILL , 6 SAMPLES **",//,
     1         "PEAK",5X,"1-1",6X,"1-2",6X,"1-3",16X ,
     1         "2-1",6X,"2-2", 6X,"2-3",16X,"3-1", 6X,
     1         "3-2", 6X,"3-3",//)
 161 FORMAT ( I3, 3F9.2,10X,3F9.2,10X,3F9.2 ,//)
 170 FORMAT (//////,8X," ** NOT ILL , 6 SAMPLES **",//,
     1         "PEAK",5X,"4-1",6X,"4-2",6X,"4-3",16X,
     1         "5-1",6X,"5-2", 6X,"5-3",16X,"6-1", 6X,
     1         "6-2", 6X,"6-3",// )
 260 FORMAT (  ////,8X,"** ILL, 6 SAMPLES**",//,"PEAK",
     1         5X,"7-1",6X,"7-2",6X,"7-3",16X,"8-1",6X,
     1         "8-2",6X,"8-3",16X,"9-1",6X,"9-2",6X,"9-3",
     1          // )
 270 FORMAT (//////,8X,"** ILL, 6 SAMPLES**",//,"PEAK",
     1          4X,"10-1",5X,"10-2",5X,"10-3",15X,"11-1",
     1         5X,"11-2",5X,"11-3",15X,"12-1",5X,"12-2",
     1         5X,"12-3",// )
      RETURN
      END
```

16.424 UBPROGRAM FOR AUTOSCALING

```
C       SUBPROGRAM FOR AUTOSCALING
        SURROUTINE AUTS (B,N,M,MEANB,SDB,ASLB )
        DOUBLE PRECISION MEANB,SDB,ASLB ,B
        DIMENSION B(2,36),ASLB(2,36),MEANB(2),SDB(2)
C
              WRITE (6,120)
              DO 93 I = 1, M
              DO 95 J = 1, N
              ASLB(I,J)=(B(I,J)-MEANB(I)) / SDB(I)
   95         CONTINUE
              WRITE (6,121) I,(ASLB(I,K),K=1,N)
   93         CONTINUE
C
  120 FORMAT (1H1,//,8X,"**AUTOSCALED PEAK RATIO*******
     1 ***************************************",//,
     1             "PK RTO" ,"1-1",4X,"1-2",4X,"1-3",4X,
     1          "2-1",4X,"2-2",4X,"2-3",4X,"3-1",4X,"3-2" ,
     1          4X,"3-3",4X,"4-1",4X,"4-2",4X,"4-3",4X,
     1          "5-1",4X,"5-2",4X,"5-3",4X,"6-1",4X,"6-2",
     1          4X,"6-3",4X,//,6X,"7-1",4X,"7-2",4X,"7-3",
     1          4X,"8-1",4X,"8-2",4X,"8-3",4X,"9-1",4X,
     1          "9-2",4X,"9-3",4X,"10-1",3X,"10-2",3X,
     1          "10-3",3X,
     1          "11-1",3X,"11-2",3X,"11-3",3X,"12-1",3X,
     1          "12-2",3X,"12-3", /// )
  121 FORMAT (I3, 18F7.2,/,3X, 18F7.2, //////)
      RETURN
      END
```

```
C      SUBROUTINE PROGRAM FOR SORTING: ASCENDING SEQUENCE
C   AND CLASSIFICATION BY K-NN METHODS
       SUBROUTINE SORT (JJ,D,M,L )
C
       DOUBLE PRECISION D
       DIMENSION JJ(10), D(26,10)
C
       WRITE (6,220)
       DO 50 K = 1,M
       L1 = L - 1
       DO 40 I = 1, L1
       IP1 = I+1
       DO 30 J = IP1, L
       IF ( D(K,I) .LE. D(K,J) ) GO TO 30
       TEMP=D(K,I)
       D(K,I)=D(K,J)
       D(K,J)=TEMP
       TEMP=JJ(I)
       JJ(I)=JJ(J)
       JJ(J)=TEMP
    30 CONTINUE
    40 CONTINUE
       WRITE (6,221) K
       DO 45 J=1,L
    45 WRITE (6,222) JJ(J), D(K,J)
       DO 46 KN =1,5,2
       WRITE (6,223) KN
       IF ( KN .EQ. 1 ) LL=1
       IF ( KN .EQ. 3 ) LL=3
       IF ( KN .EQ. 5 ) LL=5
       KNA=0.
       KNB=0.
       DO 47 J = 1,LL
       IF (JJ(J) .LE. 5 ) GO TO 10
       IF (JJ(J) .GE. 6 ) GO TO 20
    10 KNB=KNB + 1
       GO TO 47
    20 KNA=KNA + 1
    47 CONTINUE
       IF ( KNB .GT. KNA ) GO TO 11
       IF ( KNB .LT. KNA ) GO TO 12
    11 WRITE (6,224)
       GO TO 46
    12 WRITE (6,225)
    46 CONTINUE
       DO 48 I = 1,L
    48 JJ(I) = I
    50 CONTINUE
```

```
220 FORMAT (1H1,//,"** K-NN CLASSIFICATION, SORTING ****
   1 **********************************************",///)
221 FORMAT (///,"***SAMP=",I3," *****", / )
222 FORMAT (10X, 10(I4, F4.3 ), // )
224 FORMAT(5X," CLASSIFIED AS **'BEFORE INFECTION'",//)
225 FORMAT(5X," CLASSIFIED AS **'AFTER INFECTION'",//)
    RETURN
    END
```

III. RESULTS AND DISCUSSION

A. Transevaporator Sampling System

The modifications of the transevaporator technique described here, result in better overall reproducibility and eliminates the problem of the precolumn blocking during desorption due to the presence of excess water vapor.

The major modification is that a "headspace" trapping is performed before the solvent extraction step. Minor modifications are the substitution of glass beads for glass wool, the removal of the Porasil E tube and the warming of the helium during the evaporation of excess 2-chloropropane when the transevaporator is used in the extraction mode. Glass beads have been found to be a more efficient trapping material than glass wool. No differences have been found between silanized and non-silanized beads. Heating the helium carrier reduces the evaporation time for 2-chloropropane from about 20 minutes to 7 minutes. n-Pentane was substituted for 2-chloropropane in a few runs. This resulted in a lower extraction of volatiles without significantly changing the shape of the profiles.

Due to the small volumes involved, the transevaporator technique is faster than a macro headspace sampling using a Tenax trap or a micro extraction technique. Thus a complete transevaporator run with collection on both a Tenax and a glass bead tube takes about 30 minutes which is about the same time as required for a normal headspace sampling of a large sample on Tenax.

Figure 4 shows a typical chromatogram of a ten standard component mixture on 5% Witconol LA-23 stainless steel capillary column gas chromatographic system (2.6 ppb each component in distilled water). This mixture was used for testing the efficiency of the transevaporator sampling procedure.

For the technique of profile analysis to be useful, it is essential that replicate runs of the same sample yield near identical chromatograms. The reproducibility of the entire transevaporator/desorption/chromatography/integration technique was investigated by performing replicate runs of a 100 µl aqueous sample consisting of 1.3 ppb of each of the following components: n-propanol, n-butanol, 2-heptanone, benzaldehyde and 2-decanone. The transevaporator was operated in the 2-chloropropane extraction/glass bead collection tube mode. Table I lists the standard deviation and relative standard deviation for both retention time and peak area. The standard deviation for retention time varies between 5 and 10 seconds and appears to be independent of retention time. Thus retention time is an adequate parameter for defining the identity (in the context of profile analysis) of all but the earliest peaks.

In the case of peak areas, it is more meaningful to compare relative standard deviations, which, for the compounds considered, vary between 4.5% and 9.8% in a random manner. These values are acceptable for compounds in the sub ppm range and may be considered surprisingly good, if the degree of sample manipulation is considered. Our prime interest is in the profile analysis of biological fluids where components in "Normal" samples are likely to vary over a range of 20% to 50%. Thus

FIGURE 4. CHROMATOGRAM OF A TEN STANDARD COMPONENT MIXTURE ON 5% WITCONOL STAINLESS STEEL CAPILLARY COLUMN (100 m x 0.25 mm, I.D.)

1. Acetone
2. Ethanol
3. n-Propanol
4. n-Butanol
5. 2-Heptanone
6. Cyclohexanone
7. 2-Methyl-2-heptene-6-one
8. Benzaldehyde
9. 2-Decanone
10. 2-Undecanone

ORIGINAL PAGE IS
OF POOR QUALITY

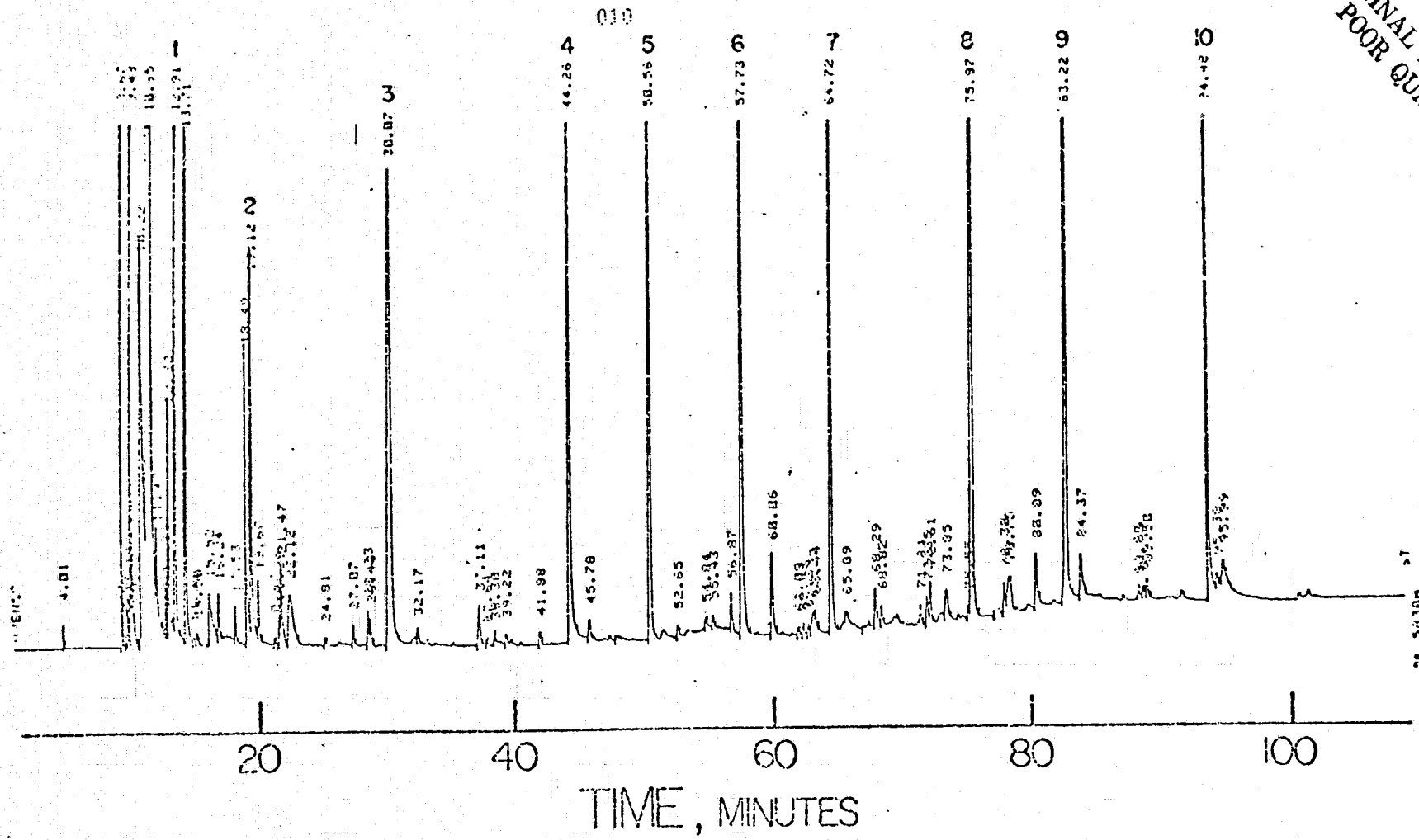


TABLE I. STANDARD DEVIATIONS OF RETENTION TIME AND PEAK AREA

Sample	Retention Time			Peak Area		
	Mean Value (min)	Standard Deviation	Relative Standard Deviation (%)	Mean Value (counts)	Standard Deviation	Relative Standard Deviation (%)
n-Propanol	26.54	0.11	0.41	9401	916	9.7
n-Butanol	36.46	0.16	0.44	37320	3228	8.7
2-Heptanone	41.81	0.08	0.19	65868	5559	8.4
Benzaldehyde	61.56	0.10	0.16	83330	3741	4.5
2-Decanone	67.40	0.08	0.12	63810	6255	9.8

changes in concentration of less than 100% are not considered significant and a relative standard deviation of 10% is more than adequate for the combined sampling and chromatographic technique. No increase in precision is obtained by considering either relative retention time or relative peak area. The reproducibility of the the technique is further illustrated in Figure 5 which shows three replicate runs of 60 µl of the same normal serum using the transevaporator in the 2-chloropropane/glass bead mode. The reproducibility is excellent, and with the exception of a few areas, the chromatograms are visually identical. The reproducibility is similar when the transevaporator is operated in the Tenax mode.

The linearity of overall response is good when the transevaporator is operated in the 2-chloropropane/glass bead mode as is shown in Figure 6 for a variety of aldehydes, alcohols and ketones. However the actual recovery (comparing aqueous sample introduced into Porasil E tube to equivalent amount of dry sample directly introduced to glass bead tube) varies from about 20% to 80% for the five compounds illustrated in Figure 7. The recovery varies slightly over a concentration range of 0.3 ppb to 5 ppb but is remarkably consistent, considering that the actual weight of each component is in the nanogram range. Because of this consistency, the non-identical recovery for different compounds is of little consequence to profile analysis which is a comparative rather than an absolute technique.

The chromatogram obtained from a given sample when using the transevaporator sampling technique depends on a number of factors such as the nature and volume of the solvent used for extraction, the volume

FIGURE 5. THREE REPLICATE CHROMATOGRAMS OF VOLATILE COMPOUNDS IN 60 µl OF NORMAL SERUM. TRANSEVAPORATOR USED IN THE GLASS BEAD MODE.

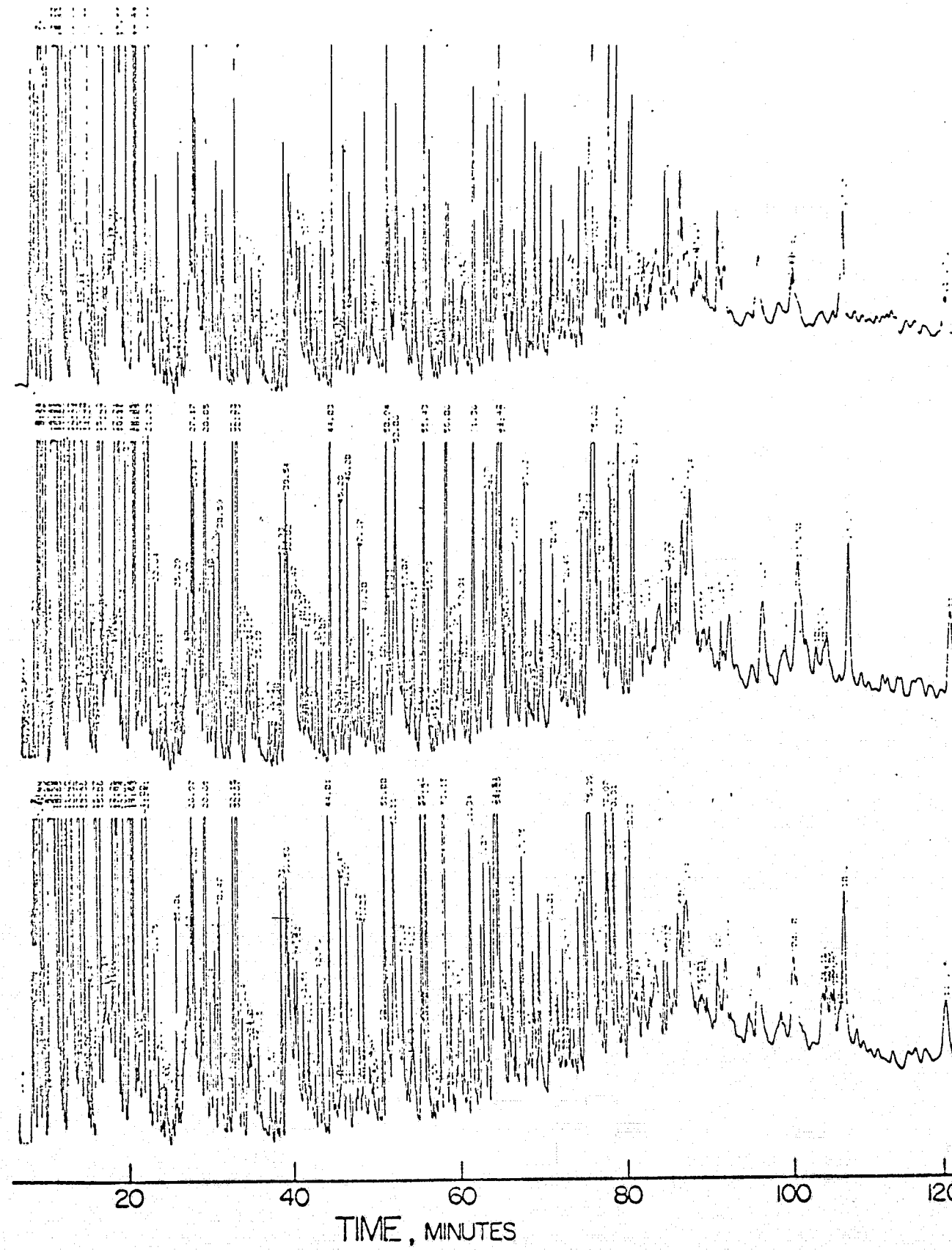

20
40
60
80
100
120
TIME, MINUTES

FIGURE 6. PLOT OF PEAK AREA VS CONCENTRATION OF VARIOUS COMPOUNDS DISSOLVED IN 100 µl OF AQUEOUS SAMPLE

1. n-Propanol
2. n-Butanol
3. Cyclohexanone
4. 2-Undecanone
5. 2-Decanone
6. 2-Heptanone
7. 2-Methyl-2-heptene-6-one
8. Benzaldehyde

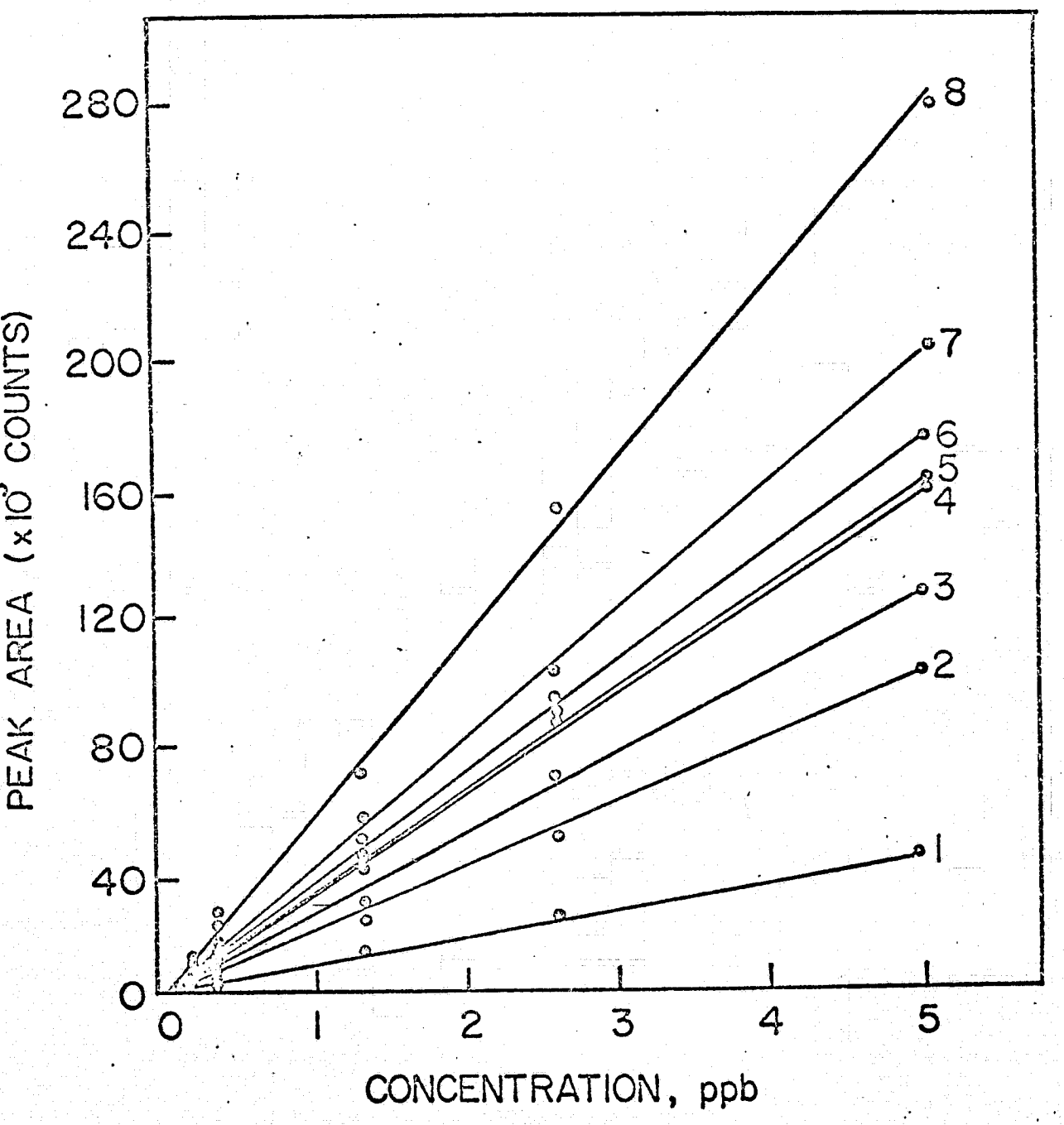
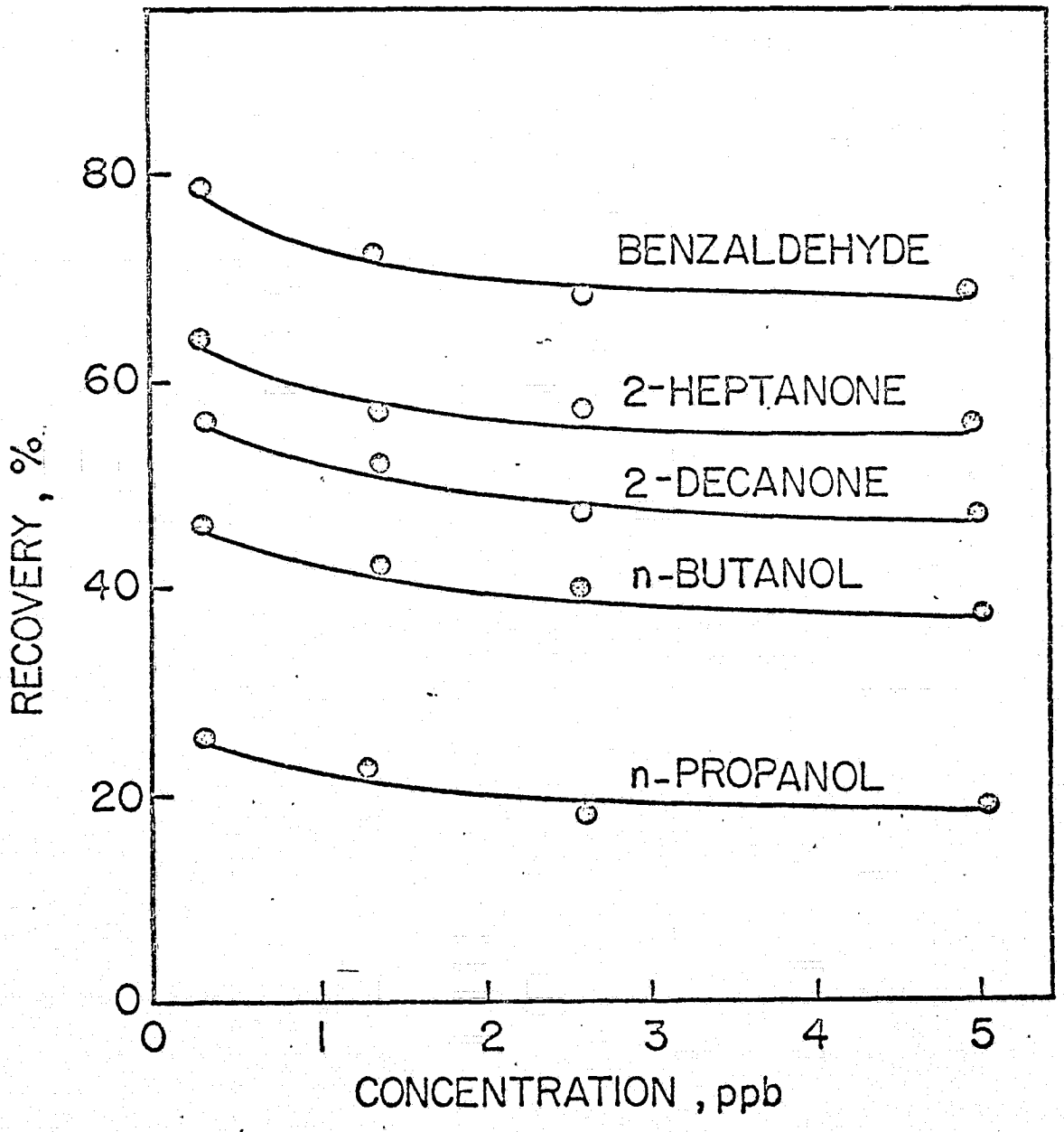

280
240
200
160
120
80
40
0
PEAK AREA (x10³ COUNTS)
0
1
2
3
4
5
CONCENTRATION, ppb
8
7
6
5
4
3
2
1

FIGURE 7. PLOT SHOWING PERCENTAGE RECOVERY VS CONCENTRATION FOR FIVE COMPOUNDS DISSOLVED IN 100 µl OF AQUEOUS SAMPLE


80
60
40
20
0
RECOVERY , %
BENZALDEHYDE
2-HEPTANONE
2-DECANONE
n-BUTANOL
n-PROPANOL
0
1
2
3
4
5
CONCENTRATION , ppb

of adsorbents and the temperature and volume of gases flowing through the apparatus during the trapping procedure. Thus a standard set of conditions must be chosen for the complete procedure when comparing profiles. A blank should be run periodically to check for contamination of the system.

The transevaporator technique has been applied to a large number of diverse samples in our laboratory. It is especially attractive for biological samples due to the small sample size requirements.

Chromatograms of normal serum volatiles (Figure 5) collected using the transevaporator in the glass bead mode were referred to earlier in connection with reproducibility studies. Figure 8 shows the corresponding chromatogram for a serum sample collected from a diabetic patient. About 200 peaks are present in each of the chromatograms. The heavy ends of the chromatograms are visually different. In addition there are several quantitative differences between the two profiles that are evident from inspection of an integrator print out. Figure 9 shows the chromatograms obtained from a diabetic serum using the transevaporator in the Tenax mode. In these profiles the relative size of the arrowed peaks is reversed in normal and diabetic serum. It is interesting that Figure 9 shows a very similar chromatogram to that of Liebich et al.[38] The latter workers used 5 ml of serum.

Differences are also found between the profiles (Figure 14) of the sera of influenza patients and the sera of controls. These differences are manifested by a greater overall concentration of volatiles in the infected sera and by ratios of peak areas that are characteristic of either the infected or non-infected sera.

FIGURE 8. A. CHROMATOGRAM OF VOLATILE COMPOUNDS IN 60 µl SERUM COLLECTED FROM A DIABETIC PATIENT. TRANSEVAPORATOR USED IN THE GLASS BEAD MODE.

B. SYSTEM BLANK

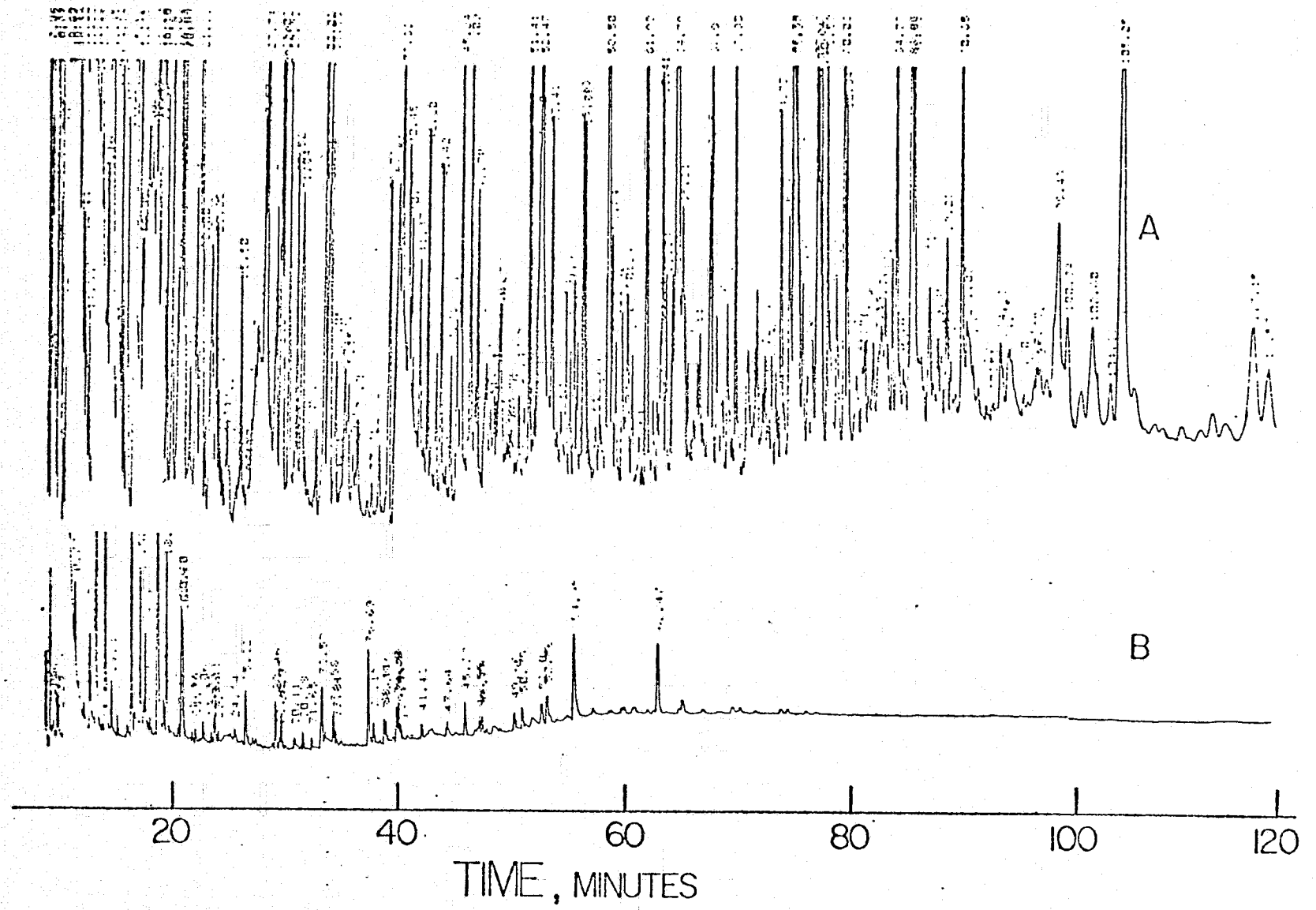
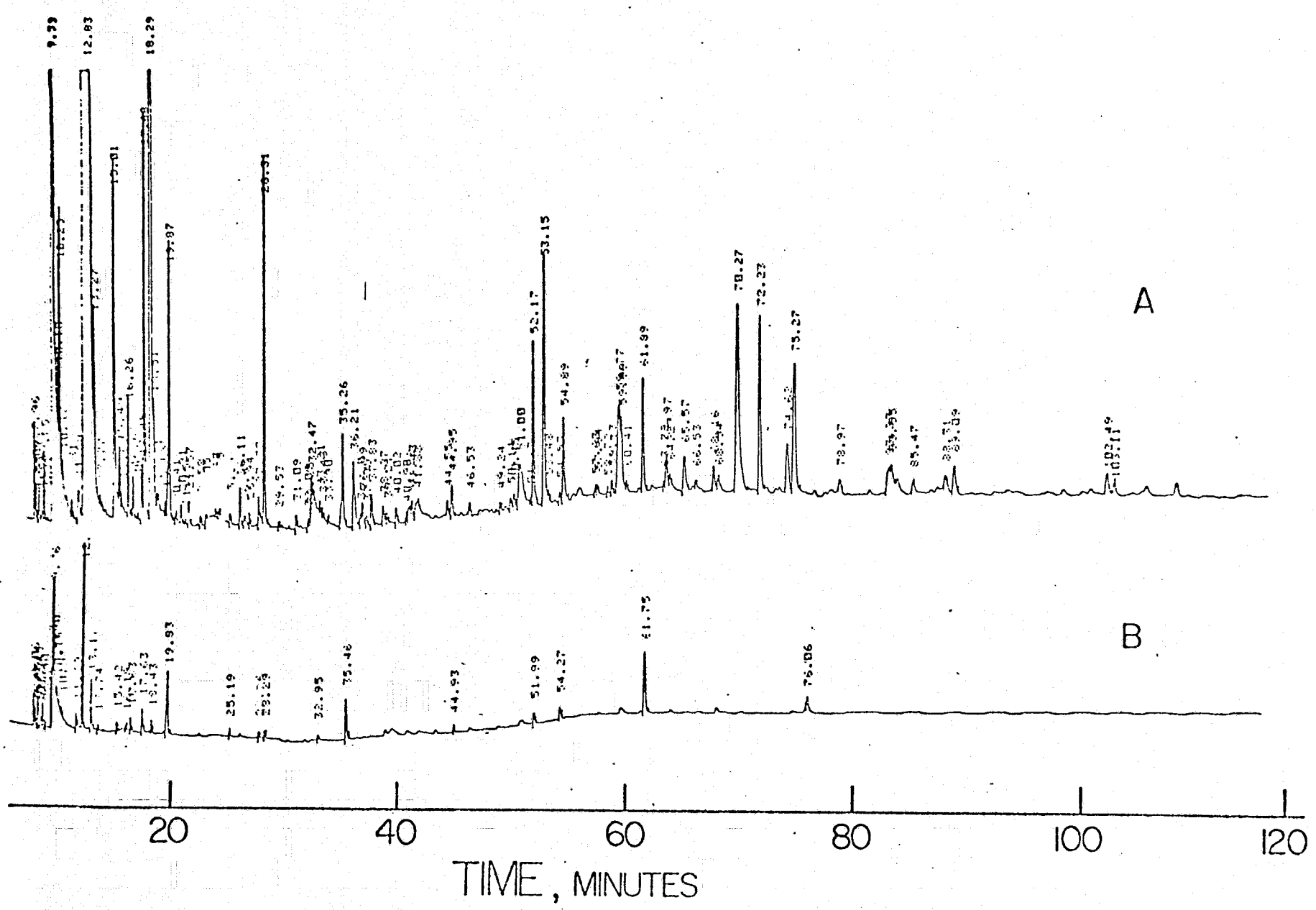

A
B
20
40
60
80
100
120
TIME, MINUTES

FIGURE 9. A. CHROMATOGRAM OF VOLATILE COMPOUNDS IN 60 µl SERUM COLLECTED FROM A DIABETIC PATIENT. TRANSEVAPORATOR USED IN THE TENAX MODE.

B. SYSTEM BLANK


A
B
20
40
60
80
100
120
TIME, MINUTES

The transevaporator technique is well suited to other body fluids apart from serum. Figure 10 shows the profile of 200 µl of saliva using the transevaporator in the 2-chloropropane extraction mode. The late peaks are small but their size could be increased four fold by increasing sensitivity without introducing noticeable noise. This profile, while not having as many volatiles as a serum profile, is of interest to biomedical studies due to the ease of obtaining samples. A profile (Figure 11) was also determined for 50 µl of breast milk which, due to its viscosity, was first diluted with an equal volume of distilled water. The profile is complex and the concentration of volatiles is larger than in serum. It is estimated that an adequate profile may be obtained with as little as 5 µl of breast milk. The profile has virtually no resemblance to commercial cow milk which has a relatively simple profile. The transevaporator technique is not suitable for urine due to the low concentration of volatiles.

Figures 12A and 12B show the different profiles obtained from samples of cognac using the transevaporator in the Tenax mode and also in the 2-chloropane/glass bead mode. Other samples that have been run by this technique include wine, cola, pressed coffee oil and brewed coffee (Figure 13), all of which give complex profiles. Differences between Robusta and Arabica coffee were found for the latter two sample types.

B. Volatile Compounds in Virus Infected Serum

A typical chromatogram obtained from a pooled virus infected serum sample is shown in Figure 14. Peaks identified in the chromatogram by

FIGURE 10. CHROMATOGRAM OF VOLATILE COMPOUNDS IN 200 µl OF SALIVA. TRANSEVAPORATOR USED IN THE GLASS BEAD MODE.

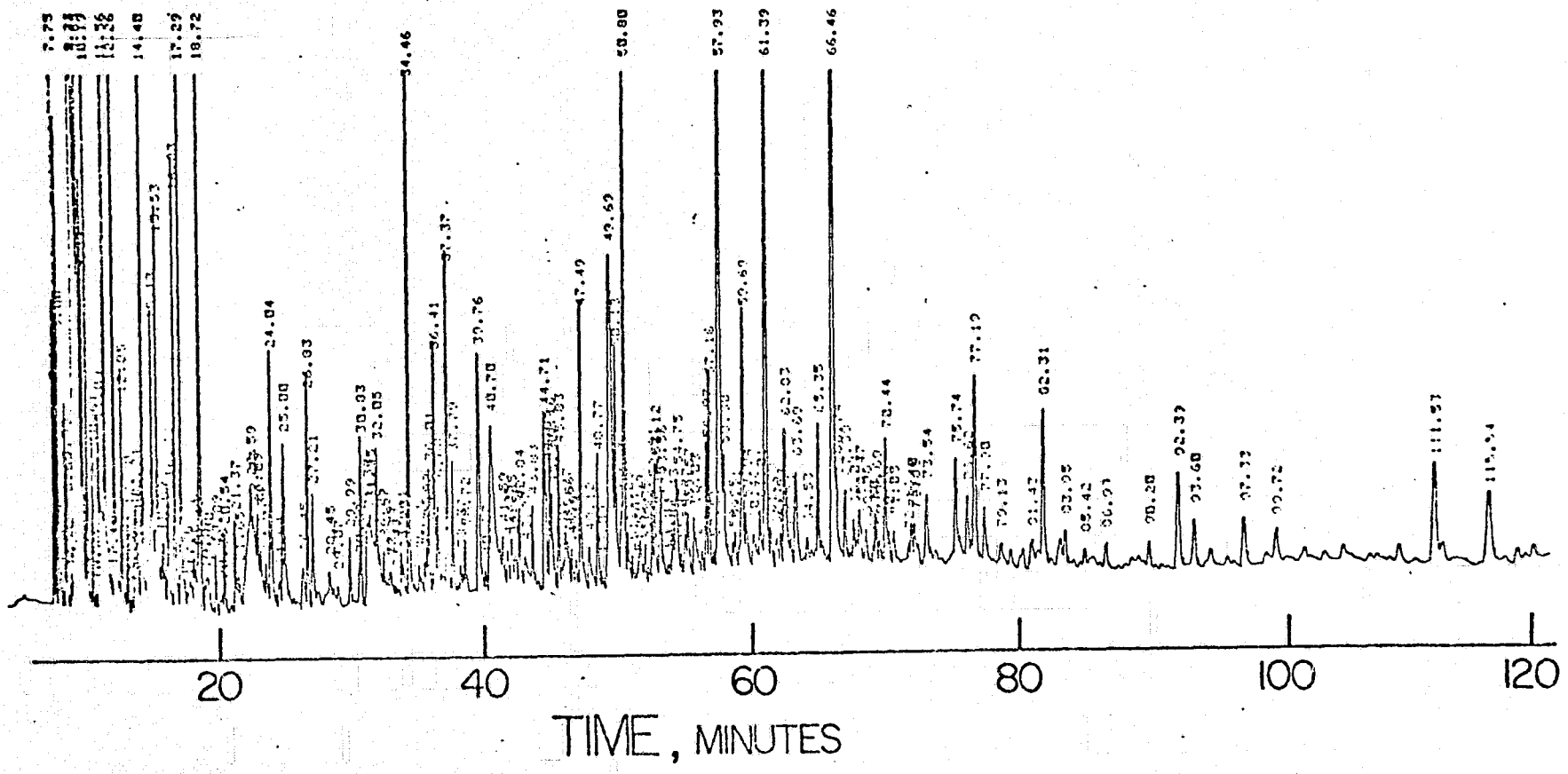

20
40
60
80
100
120
TIME, MINUTES

FIGURE 11. CHROMATOGRAM OF VOLATILE COMPOUNDS IN 50 µl OF BREAST MILK. TRANSEVAPORATOR USED IN THE GLASS BEAD MODE.

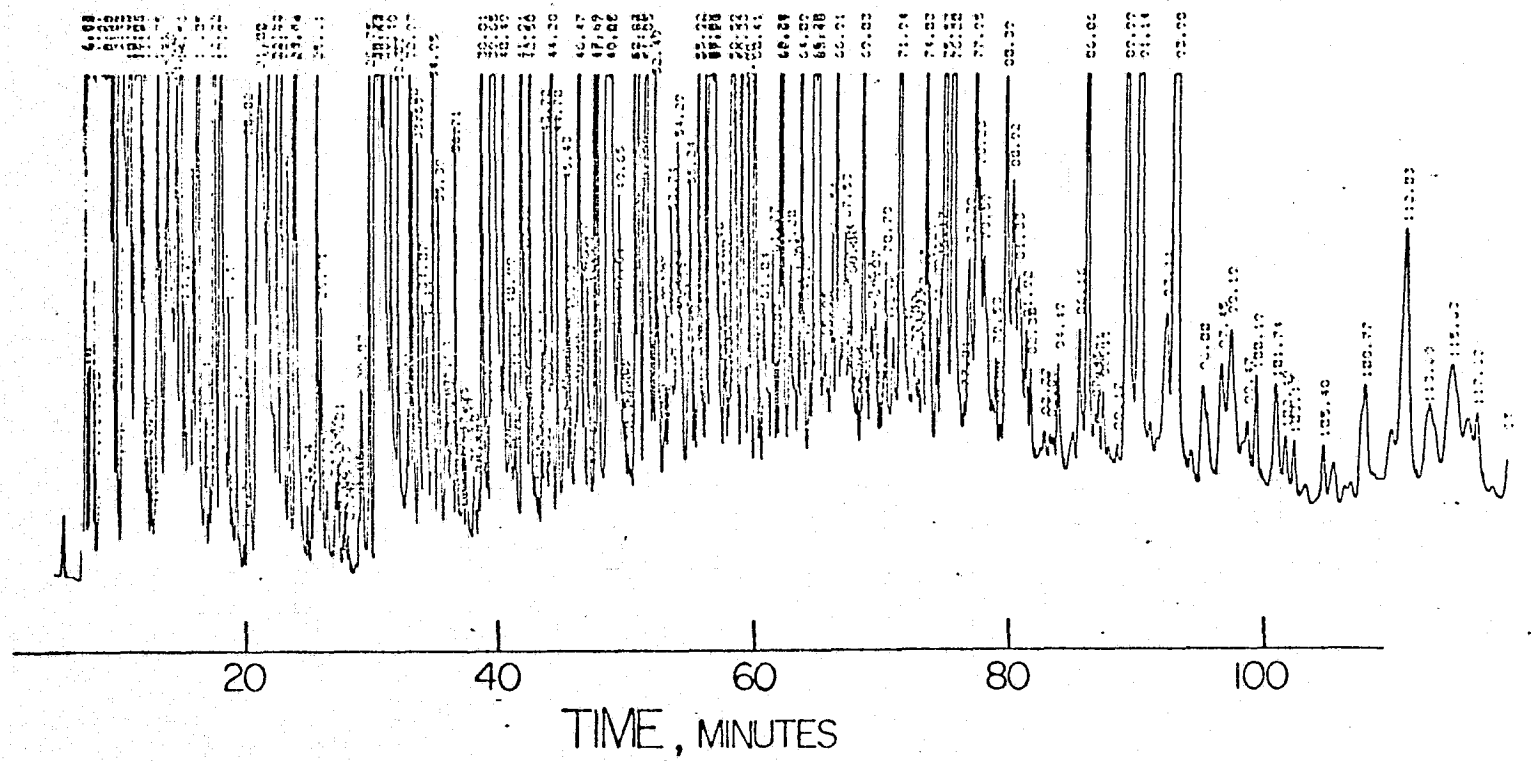

20
40
60
80
100
TIME, MINUTES

FIGURE 12. CHROMATOGRAM OF VOLATILE COMPOUNDS IN 50 µl OF COGNAC.

A: TRANSEVAPORATOR USED IN THE TENAX MODE

B: TRANSEVAPORATOR USED IN THE GLASS BEAD MODE

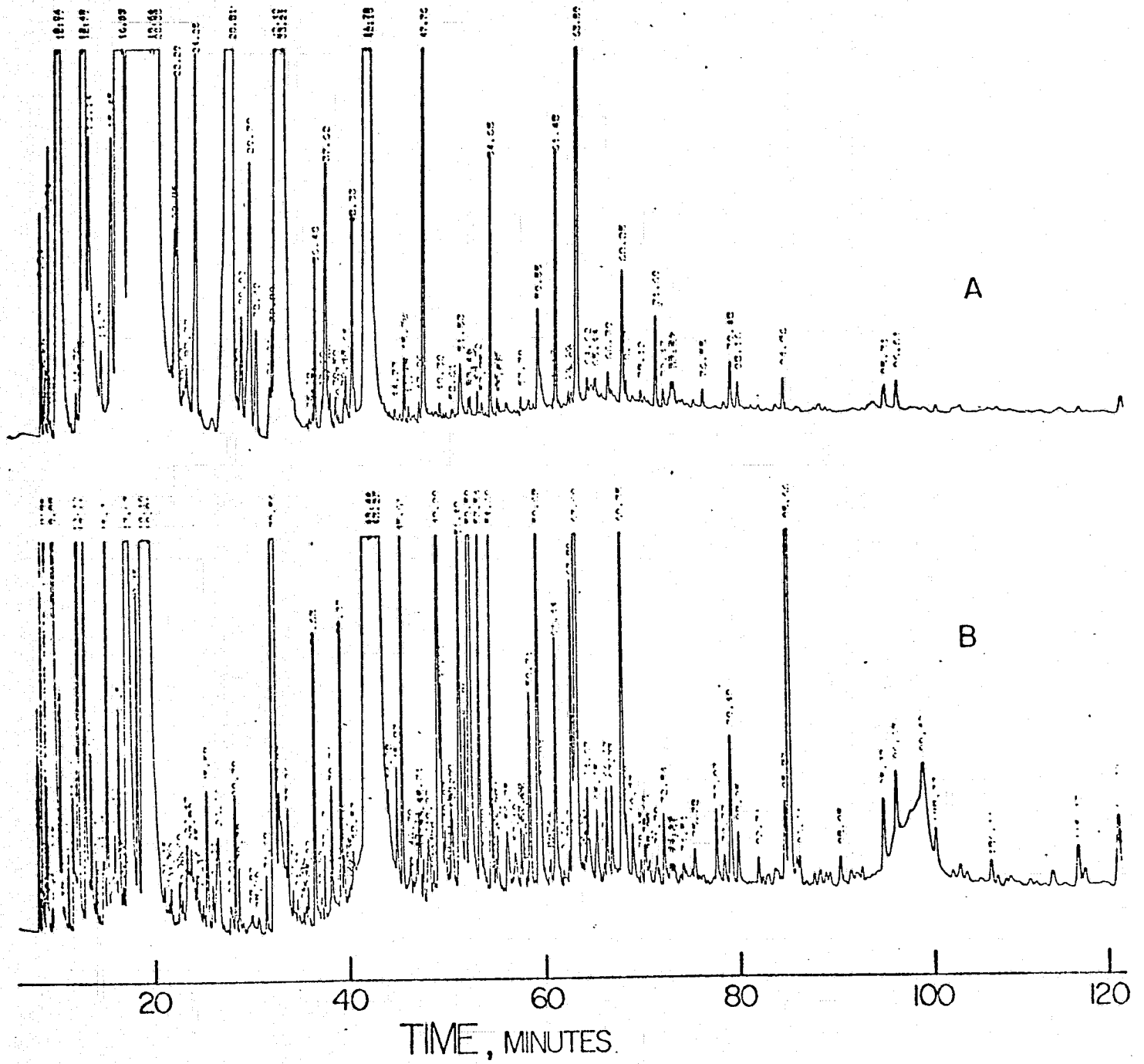
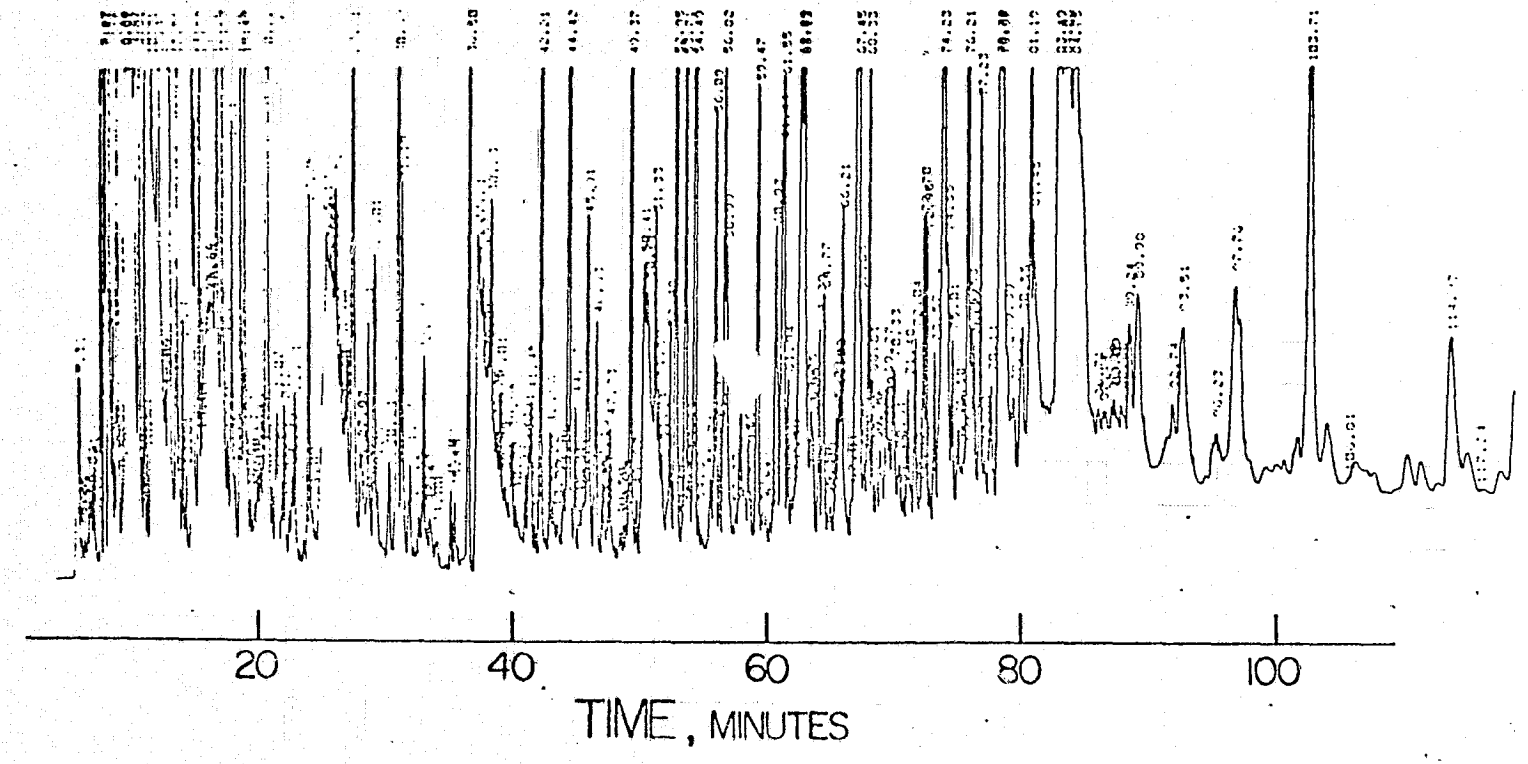

A
B
20
40
60
80
100
120
TIME, MINUTES

FIGURE 13. CHROMATOGRAM OF VOLATILE COMPOUNDS IN 100 µl OF HOT COFFEE. TRANSEVAPORATOR USED IN THE GLASS BEAD MODE.


20
40
60
80
100
TIME, MINUTES

GC-MS are summarized in Table II. The peak numbers in Table II correspond to those marked in Figure 14. We have scanned about eighty peaks in the chromatogram, but only about twenty four peaks could be identified. In the chromatogram (Figure 14) peak number 17 appears as a well defined single peak, but its mass spectrum, in Figure 15, indicates combination of two peaks; cyclohexanone and 5-methyl-3-heptanone. Five of these substances (peaks 5, 8, 10, 13, 28) have been identified previously in normal serum.[38]

C. Profile Analysis of Virus Infected Serum

A total of thirty-six virus infected and normal serum samples as well as six pooled serum samples were studied. The processing of a 70 µl serum sample including sampling by the transevaporator technique, separation by gas chromatography and data input and computation can be completed within a 2.5 hr period.

The profile of volatile substances in serum is complex with more than 150 peaks in the chromatogram. To simplify data handling, a visual comparison was made of all chromatograms, peaks due to background and stripping solvent were ignored and 37 peaks which appeared consistently in all chromatograms were selected as the data base.

It was known that the coefficient of variation for the excretion of the major and consistently excreted metabolites was large. Dietary alteration produced relatively small changes in patterns but large individual within-subject variations were observed.[30,68,69] Six replicate analyses of a pooled serum sample was used to establish the

FIGURE 14. PROFILE OF ORGANIC VOLATILES FROM A POOLED VIRUS INFECTED SERUM BY CAPILLARY COLUMN GAS CHROMATOGRAPHY.

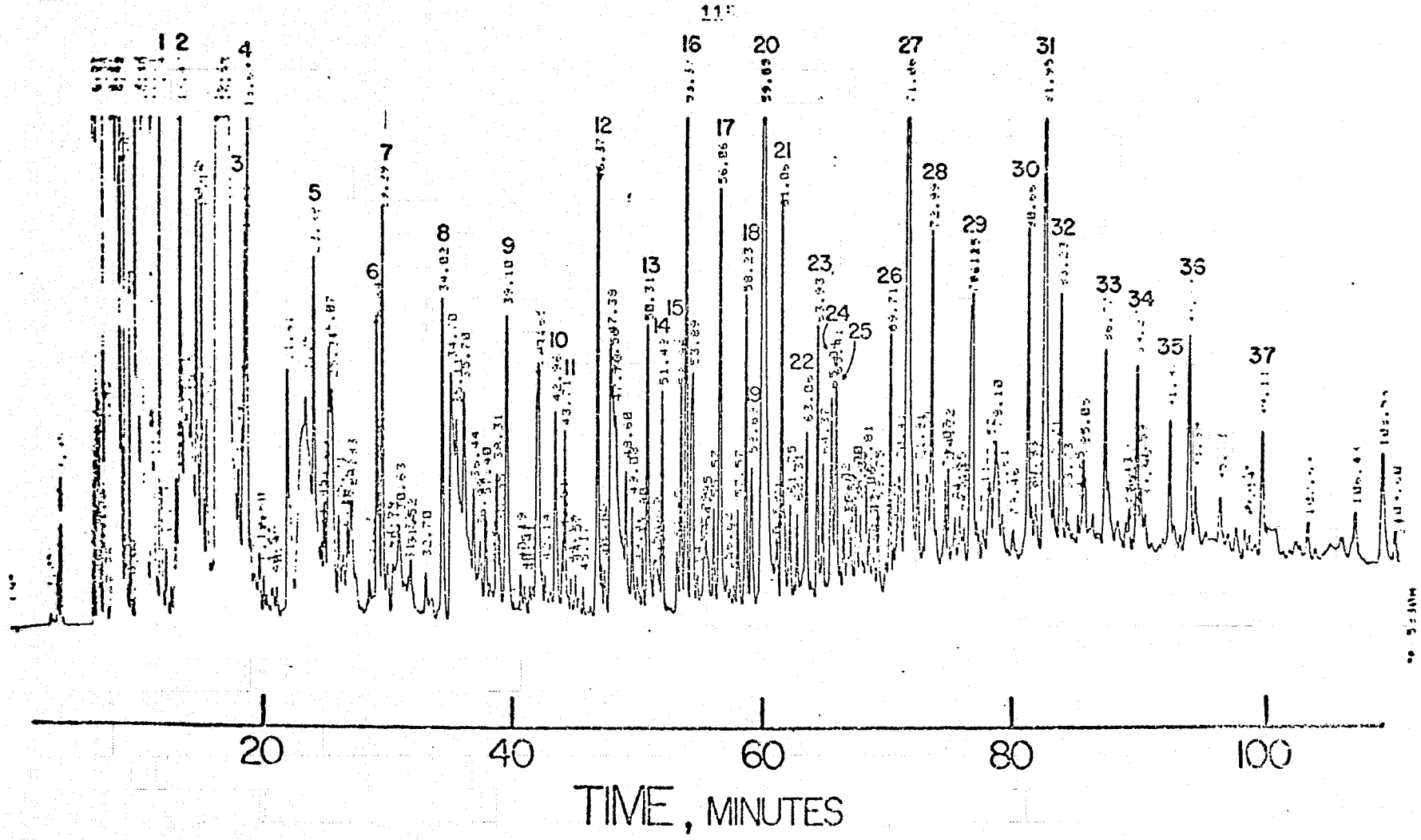

20
40
60
80
100
TIME, MINUTES

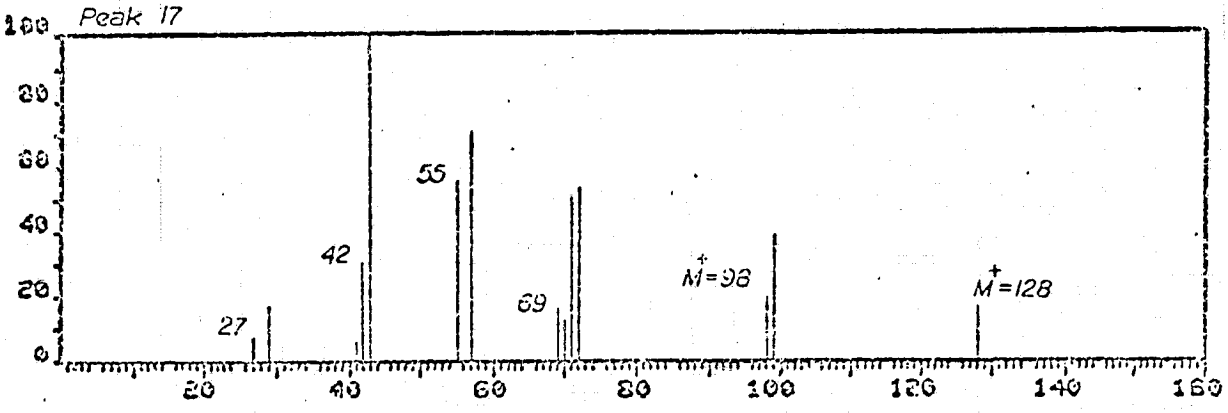
TABLE II. SUBSTANCES IDENTIFIED BY GAS CHROMATOGRAPHY/MASS SPECTROMETRY IN VIRUS INFECTED SERUM. COMPONENTS INDICATED BY NUMBERS IN FIGURE 14.

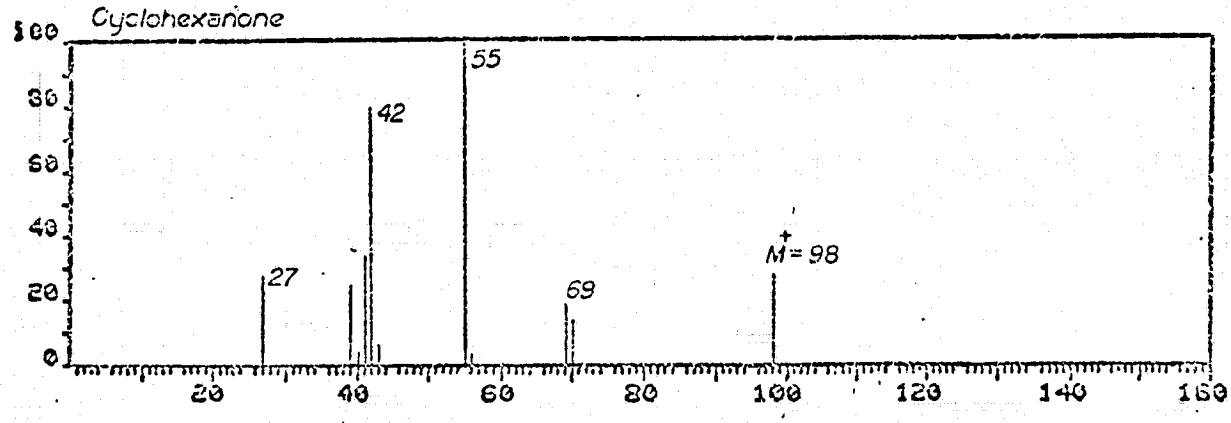
Peak No.	Compound	m/e	Molecular Weight
1	2-methyl-1-hexene	57,41,29,40,45,58,56,27	98
2	dimethylcyclopentane	56,41,55,70,42,69,39,57	98
5	2-propanol	45,43,59,31,27,29,44,46	60
7	2-hexanone	43,58,51,41,29,100,85,71	100
8	n-hexanal	44,41,43,29,45,55,72,82	100
10	n-butanol	56,31,43,41,26,27,42,57	74
12	2-heptanone	43,58,71,59,41,114,99,85	114
13	heptanal	44,70,43,55,57,42,29,41	114
14	4-heptanone	71,43,41,27,39,42,114,29	114
15	4-octanone	43,71,57,58,85,41,86,128	128
16	6-methyl-2-heptanone	43,58,71,95,85,110,70,59	98
17	cyclohexanone and	55,43,98,27,42,69,70,41	98
	5-methyl-3-heptanone	43,57,71,72,99,29,128,41	128
19	2-octanone	43,58,59,71,85,128,123,109	128
20	n-octanal	43,44,41,56,57,84,29,55	128
22	6-methyl-5-heptene-2-one	43,44,108,55,69,58,67,111	126
23	5-nonanone	51,85,58,41,142,43,42,100	142
25	2-ethylhexanal	57,72,41,43,29,56,71,55	128

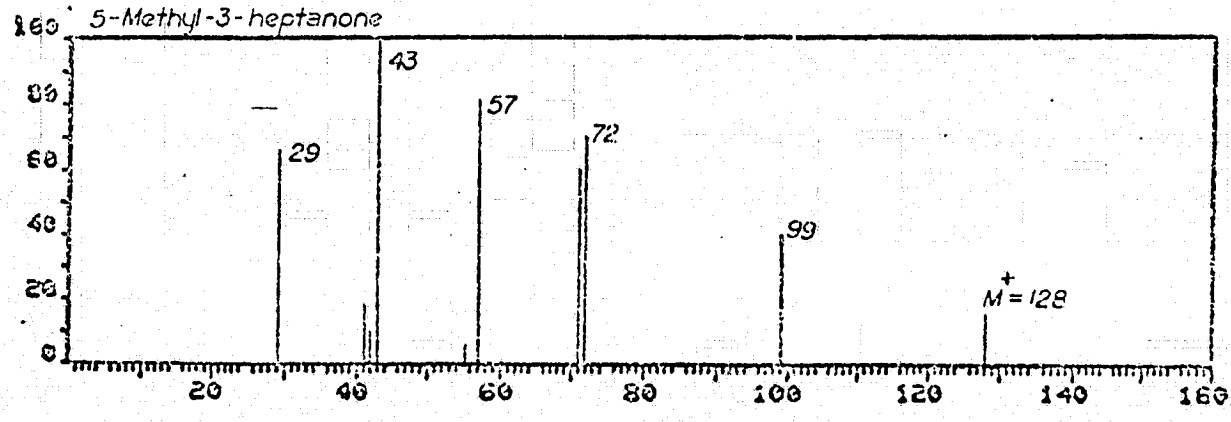
TABLE II. (continued)

Peak No.	Compound	m/e	Molecular Weight
27	2-octenol	57,56,44,43,41,29,70,82	128
28	benzaldehyde	106,105,77,41,42,40,78,43	106
29	2-ethyl-1-hexanol	57,55,41,43,46,70,83,69	130
30	2-decanone	58,43,59,71,29,57,41,55	156
32	o-tolualdehyde	120,119,91,65,82,121,51,59	120
33	acetophenone	105,77,120,51,43,78,106,50	120
34	trimethyl-2-cyclohexanone	96,138,81,109,67,95,97,123	138

FIGURE 15. MASS SPECTRA OF PEAK NUMBER 17 IN VIRUS INFECTED SERUM CHROMATOGRAM (FIG. 14), STANDARD CYCLOHEXANONE AND STANDARD 5-METHYL-3-HEPTANONE.


Peak 17
100
80
60
40
20
0
27
42
55
69
M+=98
M+=128
20
40
60
80
100
120
140
160


Cyclohexanone
100
80
60
40
20
0
27
42
55
69
M+=98
20
40
60
80
100
120
140
160


5-Methyl-3-heptanone
100
80
60
40
20
0
29
43
57
72
99
M+=128
20
40
60
80
100
120
140
160

magnitude of system variation within the data base (i.e. the effect of experimental variables on reproducibility). All peaks were normalized and the percentage relative standard deviation calculated. Reproducibility depended very much on the compound itself and consequently a wide variation was found between individual peaks reflecting their different chemical properties. The smallest variation was 6.5% (peak 20), a median value 18.3% (peak 5) and the largest 51.1% (peak 21) relative standard deviation.

Another source of variance is the variability between individuals in any one group. This was calculated for each group in turn for the 37 peaks. All peaks were normalized and their variance calculated. For example, the 12 serum samples of Group I (normal serum) had a lowest value of 17.1% (peak 1), a median value 32.9% (peak 32) and a highest value of 87.5% (peak 4) relative standard deviation. Results for Group (III) were similar but Group (II) showed far less variation.

A second data set was selected based on the magnitude of the difference between the averaged normalized peak areas for the different groups. The criteria for selection was that the relative standard deviation of the selected peaks must be greater than the variation between individuals in any one group. Seven peaks were found to meet this requirement (peaks 16, 19, 20, 25, 29, 31, 34).

The normalized peak areas in the second data set can show both negative and positive variation between groups. Under these conditions a ratio of two peaks should prove more sensitive to inter-group differences. Also, if there is an interaction between peaks in each group then peak

ratios will be more reliable for classification purposes. The seven normalized peak areas were arranged in ascending order and all possible peak ratios calculated for Groups I and II. The two peak ratios 16/25 and 29/34 were found to be most suitable for sample identification.

The reproducibility of retention time and normalized peak areas for the four selected diagnostic peaks in the pooled serum sample (six replicate analysis) is shown in Table III. Retention times can be reproduced very accurately in the analytical system and this forms an adequate parameter for peak identification. The relative standard deviation of the normalized peak areas between individuals for the four selected peaks is given in Table IV for Groups I through III. The variation in peak area for infected serum (Group II) is much less than normal serum (Group I) and can be more correctly defined. Peak 25 in Group I shows a greater variation than the others due to the inclusion of one extraordinary large peak in the data set.

1. Differentiation Between Normal Serum and Virus Infected Serum

To differentiate between normal serum (Group I) and infected serum (Group II) the two peak ratios 16/25 and 29/34 are calculated in Table V. Visual inspection shows that generally the peak ratio 16/25 usually decreased upon virus infection (except samples 7, 8, 9) and that the peak ratio 29/34 increased by virus infection (except samples 2, 3, 5). However, the use of either peak ratio does not in itself provide a sufficient classification between the two groups. The autoscaled data set of Groups I through III is plotted in two peak ratio dimensions in

TABLE III. REPRODUCIBILITY OF NORMALIZED PEAK AREA AND RETENTION TIME WITHIN A POOLED SERUM SAMPLE (6 REPLICATIONS)

Peak Number	% Normalized Peak Area			Retention time (min)		
	mean	S.D.	% C.V.*	mean	S.D.	% C.V.*
16	5.36	1.44	26.9	53.75	0.57	1.0
25	1.34	0.25	18.8	65.82	0.63	0.9
29	4.15	1.37	31.7	76.51	0.62	0.8
34	1.42	0.29	21.0	89.77	0.80	0.9

*percent relative standard deviation

TABLE IV. VARIATION OF NORMALIZED PEAK AREAS FOR THE SELECTED FOUR PEAKS IN ALL SERUM SAMPLES

Peak Number	% Relative Standard Deviation		
	Before Infection (Group I)	One Day After Infection (Group II)	14-21 Days After Infection (Group III)
16	48.9	32.6	44.2
25	89.9	20.4	47.4
29	36.3	32.7	30.5
34	53.8	30.3	62.9

TABLE V. SELECTED TWO PEAK RATIO DATA AT DIFFERENT VIRUS INFECTED CONDITIONS

Sample Number	Peak Ratio (16/25)			Peak Ratio (29/34)		
	Before Infection	One Day After Infection	14-21 Days After Infection	Before Infection	One Day After Infection	14-21 Days After Infection
Serum Samples with Clinical Symptoms After Exposure						
1	17.73	4.18	8.73	3.53	7.30	5.54
2	17.93	8.90	9.65	3.85	2.56	3.11
3	24.28	4.77	3.71	44.00	3.24	6.17
4	3.32	2.80	3.67	3.57	4.65	1.88
5	15.78	2.61	5.77	4.15	3.70	3.55
6	17.48	9.51	12.07	0.69	2.43	1.82
Serum Samples with no Clinical Symptoms After Exposure						
7	3.38	9.03	8.97	0.99	4.13	1.35
8	6.58	5.05	31.51	0.66	4.40	5.21
9	1.35	1.44	7.56	6.13	7.23	23.22
10	9.24	4.95	2.75	1.94	11.89	1.75
11	33.69	5.59	4.59	1.54	2.52	4.00
12	17.59	8.24	23.15	1.99	8.43	2.22

Figure 16. It can be seen that two separate clusters are formed for Groups I and II and thus the two peak ratios selected are adequate for the identification of normal and infected serum.

In order to test the predictability of the proposed method the K-nearest neighbor approach was used. A randomly chosen training set of ten serum samples (5 normal and 5 infected) was used to assess the predictive accuracy of the two peak ratio technique using the remaining 14 samples as unknowns. Figure 17 shows a computer print out of Euclidean distance in two factor space (16/25 peak ratio and 29/34 peak ratio). Five normal sera training set is designated as $1 \rightarrow 5$ ($B_1 \rightarrow B_{12}$) and the other five infected sera training set as $6 \rightarrow 10$ ($A_1 \rightarrow A_8$). In a given sample, the Euclidean distance between sample point and ten known points of the training set were rearranged by a sorting process. Figure 18 is a final print out for K-NN classification "before infection (normal)" and "after infection". This can be done simply by reading the first columns in Figure 18, according to already fixed K values (k = 1, 3, 5 ...etc.).

Predictive ability was calculated as the percentage of the 14 unknowns correctly classified. The three nearest neighbor (3-NN) computation assigned 85.7% of the samples correctly (1-NN, 71.4%; 5-NN, 78.5%). Clearly the proposed method has excellent selectivity for the differentiation of normal and virus infected serum.

After 14-21 days (Group III) complete recovery from infection is observed clinically. However, the two peak ratio classification places approximately half the serum samples in the virus infected category.

FIGURE 16. AUTOSCALED TWO PEAK RATIO DIAGRAM.

I. Normal Serum before infection;

II. Infected sera (24 hr);

o, Recovered sera (14-21 days after infection).

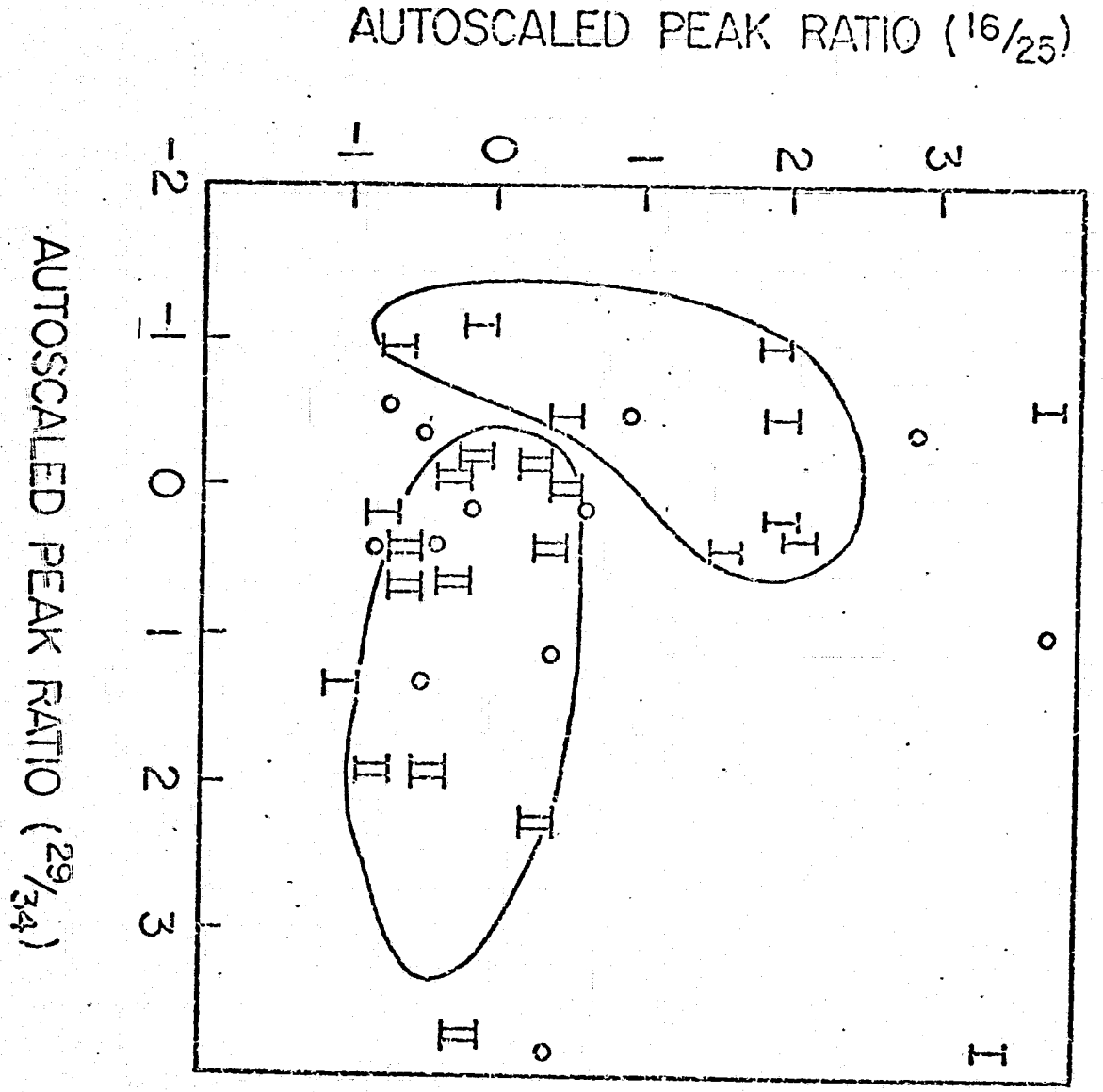

AUTOSCALED PEAK RATIO (16/25)
AUTOSCALED PEAK RATIO (29/34)
-2
-1
0
1
2
3

FIGURE 17. COMPUTER PRINT OUT FOR EUCLIDEAN DISTANCE OF 26 SAMPLES USING 10 KNOWN SAMPLES AS A TRAINING SET IN TWO PEAK RATIO FACTOR SPACE.

ORIGINAL PAGE
OF POOR QUALITY

*** K-NN CLASSIFICATION *********
******EUCLIDEAN DISTANCE D(I,J) IN 2 FACTOR **

SAMP NO	**TRAINING SET(NORMAL)** B1	B4	B7	B10	B12	TRAINING SET(INFECT) ** A1	A2	A3	A7	A8
1	0.143	2.693	2.974	1.816	0.832	2.972	1.758	2.445	1.645	2.383
2	18.290	18.636	19.780	19.173	18.981	16.961	18.905	18.731	18.200	18.209
3	0.456	2.315	2.700	1.570	1.025	2.568	1.460	2.081	1.250	1.983
4	1.281	2.916	2.603	1.620	0.590	3.858	1.769	2.616	2.199	2.834
5	2.431	1.441	0.606	0.752	2.117	3.022	0.953	1.209	1.624	1.703
6	3.248	1.208	2.351	2.388	3.530	0.750	2.134	1.445	1.681	1.041
7	3.075	5.675	5.596	4.514	2.977	6.029	4.593	5.395	4.698	5.435
8	2.805	0.490	1.654	1.704	2.978	1.228	1.470	0.726	1.173	0.434
9	2.791	0.143	1.238	1.459	2.865	1.644	1.271	0.443	1.195	0.506
10	1.600	1.253	1.304	0.226	1.502	2.408	0.125	0.955	0.775	1.210
11	3.442	1.687	2.843	2.790	3.601	0.511	2.521	1.901	1.980	1.445
12	4.448	3.762	4.929	4.559	5.032	2.075	4.273	3.900	3.579	3.380
13	2.287	0.638	0.803	0.719	2.222	2.176	0.610	0.367	0.964	0.856
14	2.819	2.372	3.476	2.936	3.372	0.907	2.653	2.429	1.945	1.913
15	1.893	1.339	2.266	1.632	2.284	1.153	1.350	1.276	0.642	0.856
16	1.502	1.169	1.508	0.536	1.543	2.143	0.287	0.912	0.475	1.029
17	2.850	1.172	2.341	2.165	3.176	0.518	1.392	1.334	1.344	0.838
18	2.703	0.762	0.413	1.030	2.570	2.442	1.015	0.642	1.414	1.160
19	2.210	0.450	1.241	0.971	2.290	1.715	0.734	0.236	0.654	0.402
20	1.302	1.793	1.645	0.522	1.023	2.864	0.667	1.494	1.181	1.735
21	1.893	1.443	1.044	0.265	1.616	2.621	0.540	1.154	1.250	1.548
22	2.651	5.252	5.530	4.368	2.951	5.127	4.339	5.020	4.179	4.894
23	9.077	8.894	10.060	9.605	9.747	7.267	9.323	9.025	8.614	8.562
24	2.883	0.827	0.370	1.203	2.742	2.515	1.195	0.765	1.578	1.265
25	2.439	0.298	1.378	1.267	2.563	1.492	1.031	0.341	0.822	0.201
26	1.161	3.719	3.693	2.573	1.035	4.183	2.634	3.431	2.748	3.481

FIGURE 18. FINAL COMPUTER PRINT OUT OF K-NN METHOD TO DIFFERENTIATE BEFORE AND AFTER INFECTION GROUPS.

```
** K-NN CLASSIFICATION, SORTING ****



***SAMP=  1 *****
              1   0.143
              5   0.832
              9   1.645
              7   1.758
              4   1.816
             10   2.383
              8   2.445
              2   2.693
              6   2.972
              3   2.974


** K-NN CLASSIFICATION ***
** K=  1
      CLASSIFIED AS **'BEFORE INFECTION'


** K-NN CLASSIFICATION ***
** K=  3
      CLASSIFIED AS **'BEFORE INFECTION'


** K-NN CLASSIFICATION ***
** K=  5
      CLASSIFIED AS **'BEFORE INFECTION'
```

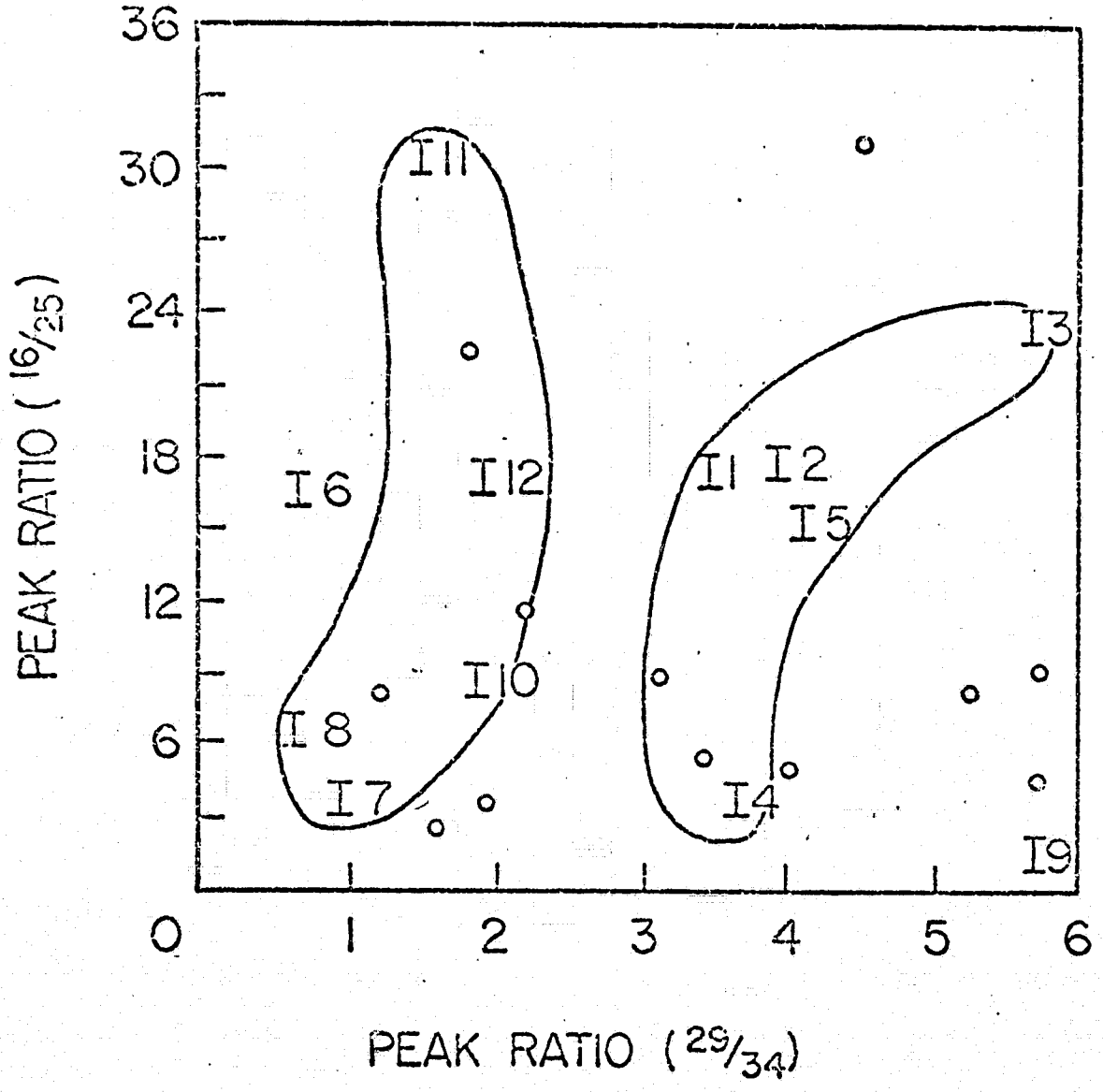
The implications of this finding can only be speculated on at the moment, especially as normal samples are correctly classified by this technique.

2. Differentiation Between Normal Serum with Clinical Symptoms After Exposure and Normal Serum with no Clinical Symptoms After Exposure

The general usefulness of this profiling technique would be enhanced if it could be used to predict susceptibility to virus infection. The normalized two peak ratio data for Group I (normal serum) samples is shown in Figure 19. Samples I1 through I6 were normal serum samples from volunteers who upon exposure to virus infection developed clinical symptoms of the disease and samples I7 through I12 did not develop clinical symptoms upon infection. Again two well defined clusters (with two exceptions I6, I9) are formed and demonstrate the possible use of the method for the diagnosis of virus susceptibility. However, the twelve samples available are too small a data base to provide a training set and sufficient unknowns to test the predictability of the method. A much larger sampling program will be required to assess the accuracy of this method for the determination of susceptibility to virus infection.

FIGURE 19. TWO PEAK RATIO DIAGRAM. SAMPLE I1 THROUGH I6 ARE NORMAL SAMPLES FROM VOLUNTEERS WHO CONTRACTED VIRUS INFECTION AFTER EXPOSURE. I7 THROUGH I12, AS ABOVE EXCEPT NO CLINICAL SYMPTOMS DEVELOPED AFTER EXPOSURE.

o, Recovered sera (14-21 days after infection).


36
30
24
18
12
6
0
1
2
3
4
5
6
PEAK RATIO (16/25)
PEAK RATIO (29/34)
I11
I6
I12
I10
I8
I7
I1
I2
I3
I5
I4
I9

IV. CONCLUSIONS

Complex profiles may be obtained from aqueous samples in the volume range from about 5 µl (for very concentrated samples) to about 200 µl by the transevaporator technique for a wide variety of biological and other sample types. Both a "headspace" and an extraction profile may be obtained from the same sample. These methods are complementary as each emphasizes different parts of the profile. The precision of the method is good considering the small sample and the trace quantities of volatiles present.

The transevaporator sampling technique is shown to be useful for the volatile profile analysis of 70 µl of influenza virus infected sera. A two peak ratio method has been developed for the characterization of normal and virus infected sera with a percentage predictability of 85.7% of correctly classified unknowns. A similar two peak ratio method is indicative as a possible means for the assessment of virus susceptibility.

The technique may be used to show differences between normal and diabetic sera, and between Arabica and Robusta coffee for both pressed coffee oil and brewed coffee.

V. BIBLIOGRAPHY

1. Biggers, R. E., Hilton, J. J., and Gianturco, M. A., J. Chromatogr. Sci., 7, 453 (1969).

2. Hawkes, S. J., and Wheaton, M., J. Food Sci., 32, 629 (1967).

3. Moskowitz, H. R., J. Agr. Food Chem., 22, 740 (1974).

4. Blumenthal, M. M., Trout, J. R., and Chang, S. S., J. Am. Oil Chem. Soc., 53, 496 (1976).

5. Dupuy, H. P., Rayner, E. T., and Wadsworth, J. I., J. Am. Oil Chem. Soc., 53, 628 (1976).

6. Bullard, R. W., and Holguim, G., J. Agr. Food Chem., 25, 99 (1977).

7. Parliment, T. H., and Scarpellino, R., J. Agr. Food Chem., 25, 97 (1977).

8. Nordby, H. E., and Nagy, S., J. Agr. Food Chem., 25, 224 (1977).

9. Palamand, S. R., and Hardwick, W. A., Tech. Quart., 6, 117 (1969).

10. Dirink, P., Schreyen, L., and Schamp, N., J. Agr. Food Chem., 25, 759 (1977).

11. Grob, K., and Grob, G., J. Chromatogr. Sci., 8, 635 (1970).

12. Nurok, D., Anderson, J., and Zlatkis, A., Chromatographia, 11, 188 (1978).

13. Bertsch, W., Chang, R. C., and Zlatkis, A., J. Chromatogr. Sci., 12, 175 (1974).

14. Overton, E. B., Bracken, J., and Laster, J. L., J. Chromatogr. Sci., 15, 169 (1977).

15. Lee, M. L., Bartle, K. D., and Novotny, M. V., Anal. Chem., 47, 540 (1975).

16. Kawahare, F. K., J. Chromatogr. Sci., 10, 629 (1972).

17. Bentz, A. P., Anal. Chem., 48, 454A (1976).

18. Flanigan, G. A., and Frame, G. M., R/D, 28(9), 28 (1977).

19. Zlatkis, A., Bertsch, W., Lichtenstein, H. A., Tishbee, A., Shunbo, F., Liebich, H. M., Coscia, A. M., and Fleischer, N., Anal. Chem., 45, 763 (1973).

20. Zlatkis, A., Lichtenstein, H. A., and Tishbee, A., Chromatographia, 6, 67 (1973).

21. Liebich, H. M., Al-Babbili, O., Zlatkis, A., and Kim, K., Clin. Chem., 21, 1294 (1975).

22. Malya, P. A. G., Wright, J. R., and Nes, W. R., J. Chromatogr. Sci., 9, 700 (1971).

23. Krotoszynski, B., Gabriel, G., O'Neill, H., and Claudio, M. P. A., J. Chromatogr. Sci., 15, 239 (1977).

24. Teranishi, R., Mon, T. R., Robinson, A. B., Gray, P., Pauling L., Anal. Chem., 44, 18 (1972).

25. Stoner, E., Cowburn, D., Craig, L. C., Anal. Chem., 47, 344 (1975).

26. Robinson, A. B., Pauling L., Clin. Chem., 20, 962 (1974).

27. Dirren, H., Robinson, A. B., and Pauling, L., Clin. Chem., 21, 1970 (1975).

28. Robinson, A. B., Dirren, H., Sheets, A., Miguel, J., and Lundgren, P. R., Exp. Gerontol., 11, 11 (1976).

29. Anbar, M., Dyer, R. L., Scolnick, M. E., Clin Chem., 22, 1503 (1976).

30. Chalmers, R. A., Healy, M. J. R., Lawson, A. M., Hart, J. T., and Watts, R. W. E., Clin. Chem., 22, 1292 (1976).

31. Malcolm, R. D., and Leonards, R., Clin. Chem., 22, 623 (1976).

32. Kitagawa, T., Smith, B. A., and Brown, E. S., Clin. Chem., 21, 735 (1975).

33. Eldjarn, L., Jellum, E., and Stokke, O., Clin. Chem., 21, 63 (1975).

34. Bultitude, F. W., and Newham, S. J., Clin. Chem., 21, 1329 (1975).

35. Goodman, S. I., Helland, P., Stokke, O., Flatmark, A., and Jellum, E., J. Chromatogr., 142, 497 (1977).

36. Jellum, E., J. Chromatogr., 143, 423 (1977).

37. Liebich, H. M., and Al-Babbili, O., J. Chromatogr., 112, 539 (1975).

38. Liebich, H. M., and Wöll, J., J. Chromatogr., 142, 505 (1977).

39. Owen, V. M. T., Ho, F. K., Mazzuchin, A., Doran, T. A., Liedgren, S., and Porter, C. J., Clin. Chem., 22, 224 (1976).

40. Martinez, M., Nurok, D., and Zlatkis, A., Submitted to J. Chromatogr.

41. Jellum, E., Størseth, P., Alexander, J., Helland, P., Stokke, O., and Teig, E., J. Chromatogr., 126, 487 (1976).

42. Lawson, A. M., Clin. Chem., 21, 803 (1975).

43. Politzer, L. R., Dowty, B. J., and Laster, J. L., Clin. Chem., 22, 1775 (1976).

44. Brown, G. K., Stokke, O., and Jellum, E., J. Chromatogr., 145, 177 (1978).

45. Heininger, J., Munthe, E., Pahle, J., and Jellum, E., J. Chromatogr., 158, 297 (1978).

46. Pfaffenberger, C. D., Malinak, L. R., and Horning, E. C., J. Chromatogr., 158, 313 (1978).

47. Zlatkis, A., and Liebich, H. M., Clin. Chem., 17, 592 (1971).

48. Matsumoto, K. E., Partridge, D. H., Robinson, A. B., Pauling, L., Flath, R. A., Mon, T. R., and Teranishi, R., J. Chromatogr., 85, 31 (1973).

49. Novotny, M., McConnell, M. L., Lee, M. L., and Farlow, R., Clin. Chem., 20, 1105 (1974).

50. Zlatkis, A., Lichtenstein, H. A., Tishbee, A., Bertsch, W., Shunbo, F., and Liebich, H. M., J. Chromatogr. Sci., 11, 299 (1973).

51. Liebich, H. M., Al-Babili, O., Huesgen, G., and Wöll, J., Z. Anal. Chem., 279, 148 (1976).

52. Liebich, H. M., and Huesgen, G., J. Chromatogr., 126, 465 (1976).

53. Dowty, B., Carlisle, D., Laseter, J. L., and Gonzales, F. M., Biomed. Mass Spectrom., 2, 142 (1975).

54. Dravnieks, A., and O'Donnell, A., J. Agr. Food Chem., 19, 1049 (1971).

55. Schultz, T. H., Flath, R. A., and Mon, T. R., J. Agr. Food Chem., 19, 1060 (1971).

56. Zlatkis, A., and Andrawes, F., J. Chromatogr., 112, 533 (1975).

57. Stafford, M., Horning, M. G., and Zlatkis, A., J. Chromatogr., 126, 495 (1976).

58. Zlatkis, A., and Kim, K., J. Chromatogr., 126, 475 (1976).

59. Isenhour, T. L., and Jurs, P. C., Anal. Chem., 43, 20A (1971).

60. Jurs, P. C., Kowalski, B. R., Isnehour, T. L., and Reilley, C. N., Anal. Chem., 41, 690 (1969).

61. Jurs, P. C., Kowalski, B. R., Isenhour, T. L., and Reilley, C. N., Anal. Chem., 42, 1387 (1970).

62. Kowalski, B. R., Jurs, P. C., Isenhour, T. L., and Reilley, C. N., Anal. Chem., 41, 695 (1969).

63. Cover, T. M., and Hart, P. E., IEEE Trans. on Info. Theory, IT-13, 21 (1967).

64. Peterson, D. W., IEEE Trans. on Info. Theory, IT-14, 26 (1970).

65. Kowalski, B. R., and Bender, C. F., Anal. Chem., 44, 1405 (1972).

66. Leary, J. J., Justice, J. B., Tsuge, S., Lowry, S. R., and Isenhour, T. L., J. Chromatogr. Sci., 11, 201 (1973).

67. Clark, M. A., and Jurs, P. C., Anal. Chem., 47, 374 (1975).

68. Lawson, A. M., Chalmers, R. A., and Watts, R. W. E., Clin. Chem., 22, 1283 (1976).

69. Chalmers, R. A., Healey, M. J. R., Lawson, A. M., and Watts, R. W. E., Clin. Chem., 22, 1288 (1976).

70. Anand, V. D., White, J. M., and Nino, H. V., Clin Chem., 21, 595 (1975).

71. Glaser, E. R., Silver, B., and Suffet, I. H., J. Chromatogr. Sci., 15, 22 (1977).

72. Liebich, H. M., and Woll, J., J. Chromatogr., 142, 505 (1977).

73. Cadogan, D. F., and Sawyer, D. T., Anal. Chem., 42, 190 (1970).

74. van Wijk, R., J. Chromatogr. Sci., 8, 418 (1970).

75. Stenhagen, E., Abrahamsson, S., and McLafferty, F. W., (Editors), Registery of Mass Spectral Data, Wiley, New York 1974.

76. Eight Peak Index of Mass Spectra, Mass Spectrometry Data Center, Aldermarston, 2nd Ed., 1974.

77. Jurs, P. C., and Isenhour, T. L., Chemical Application of Pattern Recognition, Wiley-Interscience, New York, p. 30, and p. 78, 1975.

78. Jeltes, R., J. Chromatogr. Sci., 12, 599 (1974).